1991

ECONOMIC, ECOLOGICAL, AND DECISION THEORIES

PETER A. VICTOR

JAMES J. KAY

H. JACK RUITENBEEK

Canadian Environmental Advisory Council

INDICATORS OF ECOLOGICALLY SUSTAINABLE DEVELOPMENT

Canadian

Environmental

Advisory

Council

Economic, Ecological, and Decision Theories: Indicators of Ecologically Sustainable Development

---

Enquiries concerning the work of the Council and requests for Council publications should be addressed to:

The Executive Director
Canadian Environmental Advisory Council
c/o Environment Canada
Ottawa, Canada
K1A 0H3

©Minister of Supply and Services Canada 1991
Cat. No. En92-11/1991E
ISBN 0-662-18518-8

Cette publication est disponible en français.

92-02275/

The Canadian Environmental Advisory Council

The Canadian Environmental Advisory Council (CEAC) is a body representing a cross-section of Canadians who are knowledgeable and concerned about the environment. It operates in a confidential advisory capacity to the Minister of the Environment. It provides the Minister with an alternative to the advice provided by the Department of the Environment and other federal agencies, and to the advice of specific interest groups. CEAC's public role, in terms of activities such as the publishing of reports, is therefore secondary to its primary function of advising the Minister of the Environment.

About the Authors

Peter A. Victor
VHB Research & Consulting Inc.

Dr. Victor received his Ph.D in economics from the University of British Columbia where he specialized in public economics and the economics of natural resources. His thesis, which was published as "Pollution: Economy and the Environment," was one the first major studies of the linkages between the economy and the environment. Dr. Victor's primary field of research has continued to be the economy and the environment. He has held both full-time and part-time faculty appointments at universities in Canada and England, and has done extensive work for the Government of Ontario in his chosen field. He has worked in several developing countries and is a member of several advisory boards.

James J. Kay
Environment and Resource Studies, University of Waterloo

James Kay is a professor of environment and resource studies at the University of Waterloo. He studied physics at McGill University and systems design engineering at the University of Waterloo. His Ph.D. thesis was entitled "Self-Organization in Living Systems." His principal research interest is the application of non-equilibrium thermodynamics, information theory, and systems theory in understanding the organization of ecosystems. The main purpose of this research is to develop measures of ecosystem organization. He is also involved in research concerning sustainable development and environmentally appropriate technology, particularly for housing and environmental hypersensitivity. He teaches courses in theoretical ecology, environmental engineering, and project design.

H. Jack Ruitenbeek
H.J. Ruitenbeek Resource Consulting Ltd.

Dr. Ruitenbeek is a private consulting economist, based in Ottawa, who specializes in natural resource and environmental economics. He studied at the University of Calgary and completed his Ph.D in development economics at the London School of Economics. His current interests lie in environmental planning in Third World countries, although he remains actively involved in environmental affairs in Canada.

Canadian Environmental Advisory Council
Conseil consultatif canadien de l'environnement

The Honourable Robert de Cotret
Minister of the Environment
Ottawa, Ontario
K1A 0H3

Dear Mr. de Cotret:

I am pleased to submit for your consideration the enclosed studies on economic, ecological and decision theories relevant to the ongoing work on sustainable development indicators. These studies were used to stimulate discussions at a workshop hosted by the Canadian Environmental Advisory Council in July, 1990. Due to the high quality of the authors' work, the Council decided to publish the studies as a contribution to the ongoing public demand for information related to environment-economy integration.

The Council expects to publish several reports on the subject of sustainable development indicators as a result of our investigations and consultations. We hope that the Canadian public, and those who are actively involved in work on environment-economy integration, will find our research and analysis useful, and that we will be able to provide you with practical advice on future directions and initiatives.

Please accept our best regards and our commitment to further service on your behalf.

Sincerely,

Robert Page
Chairman

Canada

116 Lisgar Street, Suite 100, Ottawa Canada K1A 0H3 (613) 943-0946 Fax. (613) 943-0940
116, rue Lisgar, Pièce 100, Ottawa Canada K1A 0H3 (613) 943-0946 Téléc. (613) 943-0940

Acknowledgements

The Canadian Environmental Advisory Council is pleased to give special recognition to the key role played by Joseph Potvin in the research, organization, and orchestration of work and events leading to the publication of this report, and to subsequent reports that CEAC will issue on the subject of indicators of ecologically sustainable development. In early 1990, Mr. Potvin was seconded full-time from the Corporate Policy Group of Environment Canada to CEAC to help guide the work on indicators. His creativity and dedication account for much of the progress over the past year.

The CEAC initiative was carried forward by a subcommittee chaired by Shirley Conover, CEAC Vice Chairman; and CEAC members Hélène Connor-Lajambe, Douglas Cook, and Peter Chapman. Ken Ogilvie, CEAC Executive Director, supported the work of the subcommittee intellectually and logistically, especially in reviewing, writing, and editing documents.

CEAC also appreciates the support received from Herman Daly and Salah el Serafy (World Bank), Peter Bartelmus (United Nations), Nancy Olewiler (Simon Fraser University), and the authors of the three delineation studies: Peter Victor, James Kay, and Jack Ruitenbeek, all of whom played key roles in the explorations leading up to the workshop and at the workshop itself. The participants, facilitators, and rapporteurs provided the broad spectrum of knowledge, experience, and conviction that so enriched the deliberations of the workshop, and CEAC extends to all its most grateful thanks.

Contents

Preface vii

Indicators of Sustainable Development: Some Lessons from Capital Theory
Peter A. Victor 1

Introduction 2
The Neoclassical School – Substitution, Technological Progress, and Markets 3
The London School – Maintaining Natural Capital Intact 8
The Post-Keynesian School – Measuring Capital 10
Natural Capital – How Useful a Concept? 12
The Thermodynamic School – Economics and the Laws of Nature 13
Conclusion 16
References 17
Appendix A: Neoclassical Theory – A Closer Look 19
Appendix B: Indicators, Models, and Constraints – An Outline of Work in Progress 21

The Concept of Ecological Integrity, Alternative Theories of Ecology, and Implications for Decision-Support Indicators
James J. Kay 23

The Development of Complex Systems 24
A Framework for Describing Responses to Environmental Change 26
Different Ecological Theories and Their State Variables 31
Some Examples of Indicators and Their Use 37
Concluding Remarks 46
References 47
Appendix A: Some Systems Notions 50
Appendix B: Ecological Stability 55

The Role of Indicators in the Decision Process
H. Jack Ruitenbeek 59

Background 60
The Decision Process: Some Issues in Political and Welfare Economics 61
Indicators as Objectives, Inputs, and Constraints 67
Indicators, Expectations, and Risk 73
Concluding Remarks 79
References 80
Appendix A: Arrow's Axioms Required of Collective Choice Processes 84
Appendix B: Attributes of Selected Approaches to Economic Analysis 85
Appendix C: Attributes of Risk Analysis Mechanisms 90

Preface

The Canadian Environmental Advisory Council (CEAC) began work on the subject of sustainable development indicators following the July 1989 meeting of leaders of the G7 summit nations. At this meeting, Prime Minister Brian Mulroney proposed that indicators should be prepared that link economics and the environment, and assess the compatibility of economic development and the protection of the environment, leading to sustainable development. The proposal was endorsed, and a recommendation was made that the Organization for Economic Cooperation and Development examine how selected environmental indicators could be developed, in the context of its work on integrating environmental and economic decision-making.

During the second half of 1989 and early 1990, CEAC reviewed studies and literature on economic and environmental indicators. Two important conclusions were reached. First, the most immediate target audience was identified as the senior policy-makers and economic analysts playing key roles in government, corporate, and development agency planning and decision-making processes. Second, since this target group relies heavily on economic analysis and resulting indicators, CEAC needed to examine how economic analysis systems could be broadened to incorporate sustaining environmental resources and life support systems. CEAC's premise was that these indicators would be used in policy- and decision-making at the highest levels of government and industry.

Since both economics and sustaining the environment were involved, the search for indicators generated discussions with leading economists and ecologists in Canada. In addition, CEAC met with staff at the United Nations and the World Bank. These investigations led CEAC to the conclusion that the development and effective use of indicators largely depends on the economic, ecological, and political paradigms subscribed to by decision-makers. Integrating economic and environmental decision-making is complicated not only by incomplete knowledge and information, but also by different theories and assumptions of economics and ecology, as well as constraints imposed by decision-making processes.

To examine the complex issues relevant to indicators of sustainable development, CEAC decided to host a two-day workshop. CEAC invited leading representatives of various schools of thought in economics and ecology together with senior Canadian policy- and decision-makers to explore the question: How do the conceptual foundations of economics and ecological analysis shape the indicators that will motivate policy in the 1990s?

To stimulate discussions at the workshop, three delineation studies were commissioned. The first two describe the major alternative schools of thought in economics and ecology, and comment on how these schools propose to deal with the new sustainable development policy agenda through concepts such as 'natural capital' and 'ecological integrity'. The authors outline some of the implications for preparing indicators of ecologically sustainable development. The third study examines decision-making processes and the further complications that they introduce into the design and choice of indicators. This third study also contains an appendix that evaluates a number of economic analysis systems for their ability to incorporate environmental factors.

In July, 1990, close to 40 experts and senior decision-makers assembled at a workshop hosted by CEAC in Ottawa. Conceptual debates centred around the three delineation studies, while more practical discussions were held on taxation, measurement, and investment. Following these discussions, the participants identified five areas of particular significance and focused discussions were held on: regenerative capacity; pricing and taxation; ethics; worst-case scenarios; and natural capital.

The July workshop was an intellectually rigorous event that stretched the parameters and perspectives on sustainable development to their

limits. The discussions were recorded in an extensive set of workshop proceedings prepared for the participants. In addition, CEAC decided to publish three documents under the general title of Indicators of Ecologically Sustainable Development:

1. Economic, Ecological, and Decision Theories
2. Synthesized Workshop Proceedings
3. Towards New Fundamentals

The first document contains the three delineation studies prepared as background papers for the workshop. The second document synthesizes the workshop proceedings, bringing together the array of ideas and insights contributed by the participants. The third document represents a distillation of the findings and sense of the workshop, as well as the work done by CEAC since July 1989 on indicators of ecologically sustainable development. It identifies what are called five new fundamentals, which should be reflected in the next generation of decision-support indicators. These fundamental principles significantly extend existing approaches to reporting environmental information, and they form part of the longer-term agenda for developing policy-relevant and operationally useful indicators.

CEAC's role in developing these indicators is first and foremost to advise the federal Minister of the Environment on the appropriate scope and direction of work in this area. While avoiding duplicating work on indicators being done by other groups, CEAC has benefited from the generous sharing of information by Environment Canada, Statistics Canada, and the National Round Table on the Environment and the Economy. The Council is not mandated to conduct ongoing research on a given subject. Hence, the three documents published represent a conceptual and, we hope, pragmatic contribution to the work being done by Canada and other nations on finding ways to integrate economic and environmental policy- and decision-making.

Indicators of Sustainable Development: Some Lessons from Capital Theory

Peter A. Victor

Abstract

A great deal can be learned by assessing the strengths and limitations of capital theory in relation to the concept of sustainable development. This paper examines the definition and role of capital in the work of selected economists who broadly represent five schools of thought in economics, and discusses some implications for the design of sustainable development indicators. Conceptualization of the environment is a major distinguishing feature of the various schools of thought. There are also differences over what is considered to be the maintenance of the stock of capital. But the meaning and practical relevance of sustainable development is closely tied to how, or indeed whether, capital theory generally applies in the context of human use of the environment.

Contents

1 Introduction 2

2 The Neoclassical School – Substitution, Technological Progress, and Markets 3

3 The London School – Maintaining Natural Capital Intact 8

4 The Post-Keynesian School – Measuring Capital 10

5 Natural Capital – How Useful a Concept? 12

6 The Thermodynamic School – Economics and the Laws of Nature 13

7 Conclusion 16

References 17

Appendix A: Neoclassical Theory – A Closer Look 19

Appendix B: Indicators, Models, and Constraints – An Outline of Work in Progress 21

VICTOR, P.A. 1991. Indicators of Sustainable Development: Some Lessons from Capital Theory. In: CANADIAN ENVIRONMENTAL ADVISORY COUNCIL. *Economic, Ecological, and Decision Theories: Indicators of Ecologically Sustainable Development.* Ottawa: CEAC. Prepared under contract for the CEAC Workshop on Indicators for Ecologically Sustainable Development (Ottawa, July 1990).

1 Introduction

This paper has been written in response to a request for a brief overview of alternative theories of economics and their approaches to sustainable development. Of particular interest is the possible relevance of each of these theories of economics for indicators of sustainable development. In response to such a daunting request, it was proposed that the paper be structured around the theory of capital and its relationship to sustainable development. At the time, this seemed in keeping with much of the contemporary literature on sustainable development, which stresses that the depletion of environmental resources in the pursuit of economic growth is akin to living off capital rather than income. The meaning and role of capital has been one of the main defining characteristics of different schools of thought in economics. Presumably, if there is much to be learned from economics about sustainable development, then discussion of capital would be key to developing our understanding.

As it turned out, this presumption was only partially correct. A review of capital theory is a useful way into an enormously complicated set of issues. Much can be learned from capital theory, in its various guises, that is relevant to sustainable development. Yet capital is not an entirely satisfactory way of conceptualizing the environment, and as much can be learned about sustainable development from the limitations of capital theory as from its strengths.

A major part of the concern about the sustainability of development relates to the environment's limited capacity to provide resources, accept wastes, and accommodate other forms of alteration, all of which are inevitable with economic activity. This is a recurring theme throughout this paper. Other aspects of sustainable development, such as a requirement for durable social and political structures and for reinforcing ideologies, are not considered.

The paper starts with an exploration of the environment as capital within the neoclassical framework. In neoclassical economics, the role of substitution and technological progress in scarcity and the effectiveness of markets in allocating resources to alternative uses at different times are central to the discussion.

The neoclassical school tends to treat all forms of capital as highly substitutable. Some economists, including Pearce and his colleagues in London, stress the differences between manufactured capital and what is frequently referred to as 'natural capital'. Pearce has proposed that maintenance of the stock of natural capital is a requirement for sustainable development. This proposal is explored under the heading of the London school. This paper suggests that such an approach leaves unresolved fundamental problems of the measurement of capital stock.

Attention then turns to the critique of the neoclassical approach to capital by the post-Keynesian school and a consideration of the relevance of this critique to the extension of capital theory to the environment. The issue of measuring what has come to be termed natural or environmental capital in some quarters is brought into focus in this discussion, since it is the measurement of capital that underlies much of the neoclassical critique of the post-Keynesian school.

At this point, the discussion takes up the issue of whether the environment is well conceptualized as capital. Perhaps capital as the 'produced means of production' of an economy does not provide an entirely sound conceptual basis for considering the environment, which is a gift of nature. According to the classical economists of the eighteenth and nineteenth centuries, land and capital play very distinctive roles in the economic process. Marshall, one of the founders of neoclassical economics, recognized this distinction but observed that economists of his day were emphasizing the similarities between land and capital rather than their differences. A review of his thoughts on this matter provides a useful commentary on the concept of natural capital.

Following Marshall's lead, it becomes apparent that any theory of the economics of sustainable development needs to be more explicit about the fundamental constraints on economic activity. Some economists, such as Ayres, Kneese, and Victor, have explored the relevance of the materials-energy balance principle for economics. This principle is based on the law of conservation of matter and energy, which is the first law of thermodynamics. Promising empirical work has been based on this approach. Others, most notably Georgescu-Roegen and Perrings, have emphasized the entropic nature of the economy, based on the second law of thermodynamics. Not only must all of the material and energy used in economic activity be disposed of back into the environment, but they are degraded and their

capacity to support future economic activity is reduced.

For at least a quarter of a century, Daly has helped develop and promote both of these sets of ideas, which are collectively referred to in this paper as the thermodynamic school. Daly has brought to the discussion an emphasis on the inevitability and desirability of a steady-state economy. Recently, he has devoted his attention to developing operational principles for sustainable development. A consideration of his ideas, especially as they relate to indicators of sustainable development and capital theory, brings this paper to a close.

Although an attempt has been made to identify the leading economists in each school, in most cases the works that are reviewed are only a small portion of a very substantial body of literature. The paper does not pretend to be comprehensive. Further, some economists have contributed to more than one school, so the assignment of an economist to any one school should not necessarily be interpreted as a rejection by that economist of the ideas associated with another school.

It should also be recognized at the outset that the range of issues considered within the context of each school is limited to a few questions that are designed to bring into focus the main ideas relating both to capital and to indicators of sustainable development.[1]

Even though some of the more technical material is confined to an appendix, much of the discussion is theoretical and demands much from the reader. However, it is hoped that at least some of the ideas presented will help establish a firmer basis for sustainable development indicators.

2 The Neoclassical School – Substitution, Technological Progress, and Markets

The interest of economists in environmental issues can be traced back to the beginnings of economics itself.[2] Francois Quesnay and other physiocrats, the first school of economics, developed their economic model around the assumption that land was the only truly productive agent or factor of production: it produced more wealth than was consumed in its creation. The physiocrats are remembered most of all for their conception of the economy as an integrated whole, but one that is tied inextricably to the productivity of nature. They even represented this graphically in the *Tableau Economique*, which stands as a distant forerunner of the modern tool of input-output analysis.

Eighteenth- and nineteenth-century writers such as Adam Smith, David Ricardo, Thomas Malthus, John Stuart Mill – all classical economists – and W. Stanley Jevons, one of the founders of neoclassical economics, concerned themselves in one way or another with the limits imposed by nature upon economic growth. In the first half of the twentieth century, the interest of economists in the environment's capacity to provide the resources needed to sustain economic activity waned considerably.[3] The 1960s saw a resurgence of interest among economists in this issue coinciding with the more general increase in environmental awareness of that period. One of the leaders of this new group of economists who were interested in environmental degradation and resources was Alan Kneese. With his colleagues at Resources for the Future, Kneese set much of the contemporary agenda in environmental economics. In their pioneering work, they drew upon the new ideas in economics that had been developed in the twentieth century largely for other purposes and which came under the umbrella of neoclassical economics.

---

[1] For example, except where it is important to those economists whose ideas are under discussion, little attention is given to the sometimes useful distinction between the various categories of resources: renewable and non-renewable, exhaustible and non-exhaustible, recyclable and non-recyclable.

[2] Economists often distinguish between the use of the environment as i) a source of mineral and biotic resources and ii) as a depository for waste products or residuals from the economic activities of production, distribution, and consumption. Much of the literature deals with one or the other of these sets of issues, so that there are separate bodies of literature on pollution and natural resources. In this paper the distinction is maintained where convenient. In other cases, the term environment is used to encompass both sets of issues simultaneously.

[3] Some important contributions to the literature were made in this period, such as those by Hotelling (1931) and Scott (1955), but these placed more emphasis on the market's ability to efficiently allocate natural resources over time, than on the implications of a failure to do so. Nevertheless, they provided useful ideas for what was to follow.

2.1 A Neoclassical Definition of Capital

An important book by Kneese and Herfindahl (1974) entitled *Economic Theory of Natural Resources* provides an erudite summary of the neoclassical approach to environmental issues.[4] One of the distinguishing features of the book is the attention given to capital theory. Herfindahl and Kneese's definition of capital is "anything which yields a flow of productive services over time and which is subject to control in production processes." They add that "this definition does not restrict capital to 'man-made durable instruments of production', a definition that enjoys the approval of some writers. Thus land is included, since it meets both of the definition's requirements. . . . As the suggested definition of capital stands, labour is a capital good." Herfindahl and Kneese recognize that the conceptualization of labour as capital may seem strange, since labourers are not bought and sold, but comment that "if it is convenient to think of some types of labour as capital, there is no bar to doing so" (1974: 68).

Writing in the early 1970s, Herfindahl and Kneese concentrated on two main roles of the environment in relation to the economy: the source of natural resources and the sink for waste products (or residuals). In later writings, Kneese recognized, along with other economists, the role of the environment in providing amenity services not usually mediated by the market, and general life support.[5]

The position of Herfindahl and Kneese, that all controllable sources of services are capital, emphasizes what natural resources and capital (as normally defined) have in common. At the same time, it implies a high degree of substitution not only among different types of manufactured capital goods, but also between these capital goods and resources. The easier it is to substitute manufactured capital for depleted resources or a degraded environment, the less concern there need be about the capacity of the environment to sustain development. For example, the substitution of satellites for intercontinental submarine cables has increased the amount of information that can be transmitted over large distances at a great saving in the use of copper. Equally, the large-scale use of air conditioning and the construction of municipal water treatment facilities has compensated to a degree for deterioration in the quality of the ambient air and water.

2.2 Resource Constraints on Economic Growth

In the period following the Second World War, economists gave considerable attention to developing models of economic growth that, in large measure, were extensions of the macro-economic ideas and concepts of John Maynard Keynes. Typically, these models used an aggregate production function for an economy in which inputs of capital and labour were combined to generate an output that could either be consumed or reinvested to increase the stock of capital. Herfindahl and Kneese made a simple extension to these models by adding natural resources to the production function and obtained two results that, if valid, are relevant to a consideration of sustainable development:

1. If the overall economy is nominally described by constant returns to scale (i.e., a doubling of all inputs will double output) and one input (resources) is fixed in supply, then the economy will exhibit decreasing returns to scale in terms of the remaining inputs (labour and capital), the amounts of which can be varied.

2. If the rate of technological progress is high enough to offset the decline in the per capita quantity of natural resource services available, output per worker can rise indefinitely.

The first of these conclusions suggests that, even if the production technologies of an economy can potentially yield increases in output commensurate with increases in inputs, overall output will be constrained by limited supplies of resources. The second conclusion suggests that these limits may be overcome by technological progress.

---

[4] Chapter 11 of the book, "Residuals and General Equilibrium: Towards a 'Spaceship' Economics" looks at the materials-energy balance approach, which has since become a more important feature of the economic treatment of environmental issues and to which Kneese also contributed. However, these ideas are not integrated into the rest of the book, and serve only as a bridge to more recent work.

[5] The distinction between amenity services and life support is somewhat blurred. The former usually refers to non-marketed services provided directly to people, for example, outdoor recreation, including aesthetic enjoyment of the environment not necessarily arising from direct use. Insofar as it is distinguished from the other three categories of service, general life support refers to the environment's role in providing habitat and the means of support for all life, however indirectly related to human existence.

Despite their contention that "natural resources are capital for the purpose of economic analysis," Herfindahl and Kneese do distinguish between resources and other capital in the analysis that led to the above results (1974: 99). They also make some provision for technological progress to increase the productivity of labour and capital to compensate for resource depletion. However, they do not explore the other avenue followed by neoclassical economists interested in the implications of a finite resource base for sustained economic growth, that is, the potential for manufactured capital to substitute for natural resources. For this, we turn to Dasgupta and Heal (1979) whose book *Economic Theory and Exhaustible Resources* is the standard neoclassical text on natural resources.[6]

2.3 Production with Exhaustible Resources

Dasgupta and Heal devote an important chapter of their book to production with exhaustible resources (Chapter 3). In particular, they say that "our concern is in the main with the implications of a finite earth for the growth possibilities open to an economy" (1979: 194). They investigate this issue with a model very similar to that used by Herfindahl and Kneese. However, their analysis is far more detailed, and accordingly, is given a much fuller treatment in what follows and in Appendix A.

Dasgupta and Heal hold population constant since "the implications of an exponentially rising population is an absurdity, if only for reasons of space" (1979: 194). They make this assumption because they are concerned with the sustainability of economic activity in the very long term. (In their analytical work they often assume an infinite time horizon.) Dasgupta and Heal also assume that there will be no technological progress to overcome problems created by a diminishing resource stock, an assumption that they then relax. Finally, they work with a production function in which some input of natural resources is essential for production, but which allows output to be maintained through the substitution of manufactured capital for resources. The result they arrive at can be summarized as follows: "even in the absence of any technological progress, exhaustible resources do not pose a fundamental problem," if reproducible capital is sufficiently substitutable for natural resources (194:205).

This result is important for two reasons. First, it is the basis for the view held by some economists that concern with the sustainability of development in a finite world is misplaced since substitution (and technological progress) can make it possible to maintain and expand the output of the economy without limit, even when the stock of natural resources is being depleted. Second, even if the optimism of these economists should prove to be unfounded, it is nevertheless important for all participants in discussions about sustainable development to appreciate the role that substitution can play in alleviating the pressures of a diminishing resource base.[7]

Appendix A offers a more detailed presentation and discussion of Dasgupta and Heal's analysis. It also draws attention to some important limitations in their conceptualization of the problem. First of all, as Dasgupta and Heal acknowledge, their mathematical formulation of how the economy can combine inputs of capital and resources to generate output (i.e., the so-called Cobb-Douglas production function) always implies a constant degree of substitutability between capital and resources, regardless of their relative proportions. In other words, no matter how far an economy goes in substituting capital for resources, the potential for additional substitution never diminishes.[8]

A second difficulty that Dasgupta and Heal also recognize is that their results depend on how capital depreciates. "It is simple to show that if . . . a fixed fraction of existing stock of fixed capital falls apart at each instant . . . the exhaustible resource is essential. Consumption must eventually fall to zero (this is assuming no technical change)" (1979: 226). Having noted this "positive embarrassment to our discussion," Dasgupta and Heal dismiss this representation of depreciation as analytically advantageous but "highly unrealistic," which seems odd, given the other equally unrealistic assumptions in their model.

---

[6] Dasgupta and Heal define a resource as exhaustible "if it is possible to find a pattern of use which makes its supply dwindle to zero" (1979: 3). This definition includes mineral resources, biotic resources such as fish and forests, and possibly, fertile land.

[7] Results very similar to those reached by Dasgupta and Heal can also be found in Solow (1974) and Stiglitz (1979).

[8] In the jargon of economics, the elasticity of substitution between capital and resources is assumed to be constant. That is, the percentage change in the ratio of capital to resources used in production brought about by a percentage change in their relative prices is constant.

A third shortcoming in Dasgupta and Heal's analysis, which is especially relevant to the concerns of this paper, is their conceptualization of resources and capital. Not only do they assume that capital does not deteriorate, but they do not recognize that capital is made from resources, whereas resources are a gift of nature. (Added to this, they explicitly assume a constant flow of service from land and "for expository ease, suppose that land is of no importance.") The trouble with treating capital and resources on an equal footing is that it ignores the fact that attempts to substitute capital for resources may be frustrated by the *increasing* demand for resources required to manufacture and operate the capital. As Appendix A shows, an extension of Dasgupta and Heal's model, in which capital itself is also produced from capital and resources, reveals the greater difficulty, even impossibility, of maintaining output in the face of limited resources through attempts to substitute capital for resources.

2.4 Neoclassical Indicators of Sustainable Development

Despite these problems, many neoclassical economists are content to rely on the kind of theoretical results generated by Dasgupta and Heal to allay any concerns about the future of economic growth stemming from a generalized scarcity of resources. Substitution and technological progress will save the day and the morrow. They are less sanguine about threats from environmental pollution. Unlike most mineral and many biotic resources, access to which is regulated by the market, the use of the environment as a depository for many types of wastes is widely recognized as being beyond the control of the market system. Here, the need for government intervention to regulate emissions or to provide economic incentives for pollution abatement is acknowledged without much argument from economists of any school.

Substitution. The analysis of production with exhaustible resources suggested that possibilities for substitution between manufactured capital and natural resources could be important in considering sustainable development. Insofar as the model of Dasgupta and Heal offers insights into these possibilities at the macro-economic level, the value of key parameters in the production function (e.g., the exponents a and b in equation 1, Appendix A) could conceivably serve as indicators of sustainable development to measure the potential role that substitution could play in the problem of a diminishing stock of resources. If a production function other than a Cobb-Douglas function is used, the elasticity of substitution between natural resources and capital could be used as an indicator.[9]

Technological Change. The neoclassical framework also emphasizes the role of technological progress as a way of overcoming environmental constraints on the economy.[10] Unfortunately, no simple, convenient measures of technological progress emerge from the neoclassical framework. In fact, it is an extremely difficult empirical problem to distinguish between pure substitution among productive inputs given the available technologies, and technological change induced by changes in the relative prices of various inputs. Nevertheless, historical measures of changes in the efficiency with which resources are used may be helpful in understanding the role that technological change has played in the past in alleviating emerging resource shortages. (For example, value added per economic sector per kilojoule of energy used could be monitored.)

If assessing the historical contribution of technological change to alleviating (or aggravating) resource scarcity is problematic, assessing its potential for the future is even more so. Apart from the uncertainty inherent in any thoughts about the future, one of the reasons that measures of the future potential of technological progress are difficult to identify is the counter-intuitive relationship that exists between expectations about technology and its likely emergence. If owners of natural resources are technological optimists, they will act in the belief that new technologies will be developed to substitute for whatever resource they currently own as that resource becomes increasingly scarce. With this view of the future, there is less incentive to conserve today. After all,

---

[9] In a Cobb-Douglas production function, the elasticity of substitution is always equal to 1. This is simply a property of this specific mathematical formulation of the production technology of the economy. Other formulations of the production function that permit the elasticity of substitution to vary may be useful for assessing prospects for substitution between resources and capital at the aggregate level. Alternatively, a more disaggregated approach could be used in which specific sectors and the technologies they employ are examined in detail.

[10] Sometimes a 'backstop' technology is merely assumed to exist. By definition, it will provide unlimited services at a fixed per unit cost that is high by today's standards, but which will limit the extent of any price increase for a resource currently being depleted. Solar energy or nuclear fusion are sometimes cited as examples of backstop technologies.

technological optimists will believe that much higher prices for resources will not materialize precisely because of the new technologies they themselves are expecting. As a result, profit-seeking resource owners with this view of the future will tend to exploit their reserves at a more rapid rate than if they believed that higher prices for their resource would not automatically induce effective mitigative technological change. Unfortunately, a high rate of resource exploitation today puts downward pressure on current resource prices, which in a world of less than perfect foresight, may undermine the very incentive for the development of new technologies on which all technological optimists are relying!

Conversely, a more pessimistic view of the likelihood of new technologies to resolve problems of resource scarcity may, through resource conservation today induced by the expectation of much higher prices tomorrow, produce short-term price signals that are far more conducive to the development of the technologies that the optimists are counting on but that the pessimists fail to see.

This sketch of the linkages between prices, expectations, and technological change highlights some of the difficulties in thinking about the contribution that technological change can make to sustainable development over the long term. In particular, it underscores the problems in devising indicators of sustainable development, especially if expectations about the future are so crucial, since reliable measures of such expectations are not only hard to invent and to implement, but also their very existence would no doubt affect the things they are intended to measure.

Prices. Neoclassical economists have done much work on the meaning and significance of resource prices, production costs, and resource rents (i.e., the difference between revenues and extraction costs). The capacity of markets to allocate a finite stock of resources efficiently over time has been analysed in great detail. One important result is known as the Hotelling Rule, which states that, in the absence of market imperfections, the net price of a depletable resource can be expected to rise at a rate equal to the rate of return on assets of the same degree of risk. This rule applies in the absence of technological progress and the discovery of lower cost supplies of the resource.[11]

Economists have examined the historical path of resource prices and extraction costs to see whether the Hotelling Rule has applied in practice. It is surmised that if substitution and technological progress are not keeping pace with increasing demands for resources, then this would show up in increases in these measures of scarcity.[12] On the other hand, if these measures have remained constant or have declined over long periods of time, then this would indicate that, to date, resource scarcity has not begun to undermine the capacity of the economy to expand.

The first major study to consider this in detail was undertaken by Barnett and Morse (1963). They examined several indicators of resource scarcity in the United States over the period 1890 to 1957 and found that, with very few exceptions, the indicators had stayed constant or declined. Their conclusion, which has since become part of the conventional wisdom of many economists and policy makers, is that substitution and technological progress have indeed been more than sufficient to overcome emerging resource scarcities. Other economists who have examined the same data as Barnett and Morse, as well as more varied and more recent data have reached different conclusions. Fisher (1981), for example, noted a decline in the rate of decrease in resource prices over the decades preceding 1970 and suggested that the effect of increasing dependence on resources was beginning to emerge. Slade (1982) went even further, and concluded that for each of 15 minerals the long-term trend in real resource prices in the United States has already begun to rise.

This is not the place to try to resolve these disagreements. A good contemporary overview of the debate about indicators of resource scarcity and their actual behaviour over time is given in Smith (1979). What is worth noting here is the possible contradiction underlying this whole endeavour. If resource markets are working according to the precepts of the neoclassical school, then resources are being allocated over time in an efficient manner. Under these circumstances, the indicators of resource prices, costs, and profits would be correct but of little practical help since, by assumption, no policy intervention is required when the market is working properly.

However, if resource markets are not operating according to the neoclassical model, possibly

[11] This rule was developed in Hotelling (1931). Dasgupta and Heal refer to it as "the fundamental principle of exhaustible resources" (Dasgupta and Heal 1979: 156).

[12] Each measure has its strengths and weaknesses. For a comprehensive review of these and others, see Brown (1979).

because of market imperfections caused by a preponderance of large companies or powerful resource-owning governments, or because the environmental effects of resource extraction are not reflected in resource prices, then the requirement for indicators becomes paramount. Unfortunately, under circumstances such as these, the indicators offered by the neoclassical school – prices, costs, profits, etc. – cannot serve since the assumptions under which they carry normative meaning do not prevail.

3 The London School – Maintaining Natural Capital Intact

Pearce and his colleagues at the London Centre for Environmental Economics have been among a leading group of economists trying to develop an appreciation of sustainable development in economic terms. The most comprehensive statement of their position is given in *Economics of Natural Resources and the Environment* (1990) but they have published similar ideas in several other places. Pezzy (1989) refers to their work as an ecological approach to sustainability. In this paper we shall refer to them as the London school.[13]

In their analysis of the sustainable economy, Pearce and Turner devote virtually all their attention to the meaning and desirability of maintaining the natural capital stock as a condition for sustainable development (Pearce and Turner 1990, Chapter 3). They do not define this term, though they suggest it is synonymous with the stock of environmental resources (1990: 46). Further, they offer three reasons for differentiating between natural capital and manufactured capital,[14] and in so doing, distance themselves from the position of Herfindahl and Kneese. These reasons are:

1. Manufactured capital is not independent of natural capital. The latter is often needed to make the former.
2. Natural capital fulfils other economic functions, including basic life support; it is multifunctional to an extent not shared by manufactured capital.
3. Following from the above, it is not always possible to substitute manufactured capital for natural capital.

Pearce and Turner develop their position on maintaining the stock of natural capital by pointing out that "new technology is not necessarily less polluting" (1990: 50) and they question the long-term prospects for resource-saving technological progress. In particular, they challenge the oft-made assumption in the neoclassical literature of backstop technologies, which are assumed to provide an upper limit to the price of resources. They specifically question the assumption underlying any backstop technology, which is the existence of an "indefinitely renewable resource which eventually takes over when exhaustible resources have gone."

3.1 Uncertainty and Irreversibility

Pearce and Turner consider two aspects of natural environments important in any discussion of sustainable development: uncertainty and irreversibility. There is a great deal of uncertainty about the role that the natural environment plays in supporting economic activity and about the effects of economic activity on the environment. As Perrings (1987) has pointed out, this is real uncertainty in Knight's sense that probabilities cannot be assigned to alternative outcomes. The reason for this is that the set of possible outcomes is often unknown, and without this information a probabilistic approach is untenable.

Following Krutilla (1967), Pearce and Turner note that changes in the natural environment caused by human activities may be irreversible to a degree not matched by changes in manufactured capital. For example, buildings can be constructed and later demolished. Equipment can be manufactured and later dismantled and sold as scrap. However, once an oil deposit has been depleted and the oil burned for energy, or once an area of wilderness has been harvested for trees, then it is effectively impossible to recreate these components of the natural environment. Pearce and Turner conclude, therefore, that: "the presence of uncertainty and irreversibility together should make us more circumspect about giving up natural capital" (1990: 51).

---

[13] It may seem premature to credit Pearce et al. with founding a new school, but there can be no doubt that their ideas have already become influential in the profession and in government, and promise to become more so in the years ahead.

[14] They use the term man-made capital, but manufactured capital seems more appropriate.

With three more points, Pearce and Turner conclude their argument for maintaining the stock of natural capital (1990: 51-52). These can be summarized quickly:

1. In some circumstances, such as a rural setting in a developing country, more natural capital can mean more resilience to shocks, and hence, a more sustainable society.

2. Considerations of intergenerational equity demand that the resource stock be maintained so as to ensure broadly equal access to it by different generations.

3. Preservation of natural capital is consistent with a world view that recognizes the rights of other species to co-exist with humans.

3.2 Indicators of a Constant Stock of Natural Capital

Having put forward numerous reasons for maintaining the stock of natural capital as a condition for sustainable development, Pearce and Turner turn their attention to the meaning of a constant stock of natural capital.[15] It is worth quoting them at length as they attempt to explain what a requirement that the natural capital stock be constant might mean:

"There are several interpretations. First, we could say that the capital stock is constant if its *physical quantity* does not change. But we have no way of adding up the different physical quantities (tonnes of coal, cubic metres of wood, litres of water, etc.). The standard economic approach would be to value each type of resource in money terms and compute the overall aggregate money value. If this could be done, in the same way as we make estimates of the 'national wealth' – i.e., the stock of man-made capital – then we could rephrase the natural capital requirement in terms of a constant real value of natural assets.

"Second, we could think in terms of the unit value of the *services* of natural capital. That is, we could look at the *prices* of natural resources and aim to keep these constant in real terms. Provided we are satisfied that prices reflect absolute scarcity . . . constant real prices will imply a constant natural capital stock in this modified sense. One obvious problem here is that many resources do not have observable prices. We would need to find implicit or 'shadow' prices in some way.

"Third, we could think of a *constant value* of the resource flows from the natural capital stock. This is different from constant prices because we would allow quantity to decline but the price to rise, keeping value constant" (1990: 53).

To summarize, in their three points Pearce and Turner offer four interpretations of the requirement to hold constant the stock of natural capital:

1. The *physical quantity* of natural resource stocks should remain unchanged.

2. The *total value* of the natural resource stocks should remain constant in real terms.

3. The *unit value of the services* of the natural resources, as measured by the prices of natural resources, should remain constant in real terms.

4. The *value of the resource flows* from the natural resource stock should remain constant in real terms (where resource flow is the product of price and quantity used).

Pearce and Turner recognize some of the shortcomings of each of these definitions of a constant stock of natural capital. Measurements of natural capital stock made exclusively in physical terms are problematic because of the difficulty of adding up different physical quantities expressed in different units. If the standing stock of timber increases at the same time as the deposits of natural gas decrease, how can it be determined whether the stock of natural capital has risen, fallen, or stayed the same?

It is for this reason that the second interpretation of a constant stock of natural capital is offered. By valuing each resource stock in monetary terms, the total value of natural capital can be measured. The prices, or prices net of extraction costs, of the resources could be used to value the capital stock, but there are several problems

[15] They also consider whether the stock to be maintained should be the existing stock or the optimal stock and conclude that, as a strategy for dealing with uncertainty and irreversibility, the existing stock at least should be maintained or, in the case of poor countries, increased.

with this approach. First, there are no market prices for resources such as air, water, or wilderness, and shadow prices would have to be estimated. Second, those prices that do exist may not be useful: they may be affected by market imperfections and taxes, and they may exclude externalities involved with the production and use of the resource. Further, they do not adequately capture the interests of future generations. In other words, they may have little or no normative relevance for valuing the stock of natural capital.

There are additional problems in using market prices to value the stock of resources with a view to keeping the aggregate value of resource stocks constant as a condition and indicator of sustainability. Resource prices or net prices reflect conditions at the margin and to use these to value entire stocks is a highly questionable procedure which can give perverse results. For example, it is possible for the real price or net price of a resource to rise over time at the same rate as (or faster than) the rate of decrease in the physical stock of the resource. This will result in the value of the resource remaining constant (or rising) until there is none left, which is hardly consistent with sustainability, given Pearce and Turner's own rejection of backstop technologies as a way out.

This possibility is of more than theoretical interest. If price or net price rises as resource quantity is declining, the value of resource stocks as an indicator of sustainability can give precisely the wrong policy signal to government. As long as the value of the stock remains constant or rises, the government will not perceive a problem through this indicator, even though the flow of resources is becoming increasingly valuable (as measured by price) and the physical stock is declining. Conversely, government intervention might well be prompted if the stock of resources is declining, even though this is because the flow of resources, again measured by its price, is becoming less valuable.

Pearce and Turner's third and fourth interpretations of a constant stock of natural capital also use market prices, and so similar criticisms made in relation to keeping the value of the capital stock constant apply and need not be repeated. In light of all these difficulties, it seems that the value of resources or resource flows could be a very unreliable indicator of sustainable development. Use of such an indicator might lead to policy initiatives that could exacerbate the problems it is intended to forewarn.

For most of their discussion of measuring the constancy of the stock of natural capital, Pearce and Turner abstract entirely from the possibility of substituting manufactured capital for natural capital, which was the mainstay of the neoclassical approach. Conceivably, Pearce and Turner's position could be restated in terms of the *combined* stock of natural and manufactured capital, since it is from both of these that current and future generations derive their livelihood. Although they do not go this far, they do say that "sustainability can be analysed in terms of a requirement to maintain the natural capital stock . . . how far it is possible to relax this requirement depends on what we believe about the degree of substitutability between renewable and exhaustible resources, and between man-made capital and natural capital. It also depends on the behaviour of technological progress in reducing the resource input to a unit gain in the standard of living" (1990: 52).

In view of the importance given by Pearce and Turner to the maintenance of the natural capital stock as an indicator of sustainable development, and the obvious promise this important idea holds, it is unfortunate that they did not devote more attention to the theoretical and practical problems of measuring the capital stock. These problems are unlikely to be overcome easily. Indeed, the problem of measuring capital has been one of the fundamental sources of criticism of neoclassical economics levelled by the post-Keynesian school. Although members of this school have paid no attention to the problems of economy-environment dependencies, lessons can be learned from their analysis of (manufactured) capital, since their misgivings about the neoclassical approach to capital also apply to natural capital.

4 The Post-Keynesian School – Measuring Capital

In the discussion on neoclassical perspectives in Section 2, an aggregate production function was used to arrive at some conclusions about substitution and continued economic growth, which if valid, are highly relevant to a consideration of sustainable development. One of the variables in the production function was capital. This variable is supposed to represent the sum total of capital goods in the economy. In the neoclassical framework there is a marginal product of capital that determines, in conjunction with all of the other relevant marginal products and marginal utilities, the prices prevailing in the

economy and the distribution of incomes among the owners of land, labour, and capital, the so-called factors of production.

According to Joan Robinson and her followers, this neoclassical theory assumes that the quantity of capital can be measured independently of its price. After all, the distinction between prices and quantities is fundamental in the neoclassical analysis of markets and it is the interaction of multiple markets and a general equilibrium in those markets that is the focus of attention in neoclassical economics.

The neoclassical models of production and growth reviewed in Section 2 use an aggregate measure of capital (K in equations 1 and 2 of Appendix A). Price is the common denominator used to add up such things as buildings, tractors, lathes, and the entire panoply of capital goods into a single value of capital expressed in monetary units and used in such models as these. One can see immediately the problem in trying to measure the quantity of capital independently of the price of capital, if the prices of individual capital goods are used to combine different items of capital into an aggregate quantity of capital.

There is a further aspect of the capital measurement problem that again highlights the difficulty of distinguishing between the quantity of capital and its price. The price of a capital good reflects its value to the buyer. This value is derived from the increase in output over time that is made possible by the acquisition of the capital good. To assess the present value of this increased output, and therefore, the maximum price the buyer will pay for the capital good, an interest rate must be used. According to neoclassical economics, the interest rate is the price of capital determined in the capital market. Hence, we are again faced with a circularity: the quantity of capital depends on its value, its value depends on the rate of interest, the rate of interest depends on the quantity of capital.[16]

There are many aspects to the debate between the neoclassical economists and the post-Keynesians that are beyond the scope of this paper.[17] What is central to the theme of this paper is the difficulty of finding a satisfactory measure of aggregate capital. The post-Keynesians have addressed themselves only to manufactured capital, but their arguments apply with as much force to natural capital. The London school, which has placed so much emphasis on maintaining the stock of natural capital, has not faced these issues.[18] Moreover, the difficulties involved in finding theoretically sound, robust measures of the stock of natural capital may be even greater than those identified by the post-Keynesians for manufactured capital. For example, *known* stocks of nonrenewable resources, including recyclables, depend on current and expected future resource prices since these provide the incentive for exploration. The value of these resource stocks depends on their physical magnitude, current and future expected prices and extraction costs, and the rate of interest. Future prices depend on future demands, which depend on future incomes. These future incomes depend, in part, on future stocks of natural and manufactured capital which depend, in turn, on decisions taken today to exploit or conserve the known stocks of resources and to invest in manufactured capital.

This is an enormously complicated, interconnected system that seems to belie any clear distinction between the quantity of capital and the price of capital. Thus, even if the objective of maintaining the stock of capital is accepted as a condition of sustainable development, the problem of developing relevant indicators remains.

What has been said here about the post-Keynesian school of economics is really only the stepping-off point for what has turned into a far-

---

[16] In defence of their position, some neoclassical economists have argued that this is an issue of simultaneity rather than circularity. In other words, economists have always recognized that myriad prices and quantities in a market economy are interrelated and determined through the mutual interactions among numerous buyers and sellers.

[17] See Hamouda and Harcourt (1988) for an overview of the post-Keynesian position on economic growth, distribution, and other issues in macro-economics. For a contrary view of the overall significance of the 'Cambridge Controversy' see Burmeister (1980), though even he concedes that "it is evident that theoretically rigorous aggregation of heterogeneous machines into a single, well-behaved index of 'capital' is possible only under extraordinarily restrictive assumptions" (1980: 154).

[18] Neither has the neoclassical school. For example, in Herfindahl and Kneese (1974: 97): "If natural resources are incorporated in the production function as are (manufactured) capital and labour, the implicit assumption is that some unit has been found in which the quantity of natural resources entering into production can be expressed. To do this in application necessarily would involve an index number problem of the same degree of complexity as a similar attempt for capital. *Waiving this problem* . . . (emphasis added).

reaching critique of neoclassical economics that rejects the marginal productivity theory of income distribution.[19] In its place, post-Keynesians have offered a theory of income distribution in which capital no longer plays "the dual role it does in the neoclassical theory: i) as a factor of production, and ii) as the recipient of the distributive share called profit" (Ahmad, unpublished manuscript). Only the second of these roles is allowed, and depending on the specific formulation, the relative shares of capital (received by capitalists) and of labour (received by workers) are determined by such things as the capitalist's desire to invest and the savings rates of the different socio-economic classes.

Although this is not the place to explore these issues further, insofar as the distribution of income, wealth, and access to the environment is relevant to sustainable development, the various schools of thought in economics differ as much on this issue as they do on the meaning of capital itself.

5 Natural Capital – How Useful a Concept?

At various points in this paper, attention has been drawn to the fact that manufactured capital is used in production and itself is produced by the economy, that is, it is a produced means of production. The environment is not produced, though it can certainly be altered and made more or less productive in economic terms by human actions. Some economists, especially those in the neoclassical school, tend to play down this distinction and some (Herfindahl and Kneese, for example) ignore it all together. But if the distinction between manufactured capital and natural capital is important, then it calls into question the conceptualization of the environment as capital in the first place.

Interesting ideas on this issue were put forward by Alfred Marshall, one of the most influential economists of the first part of this century. As has been emphasized by Joseph Potvin (1990: 15-16), Marshall drew a distinction between capital and land, "the latter being natural agents . . . taken to include all free gifts of nature, such as mines, fisheries, etc., which yield income" (Marshall 1920: Chapter 4).[20] He argued that "land is but a particular form of capital *from the point of view of the individual producer*" (emphasis added), and that farmers will try to combine all their resources, including land, in a way that maximizes their profits. He goes on to say that "there is this difference from the point of view of society. If one person has possession of a farm, there is less land for others to have. His use of it is not in addition to, but in lieu of the use of a farm by other people; whereas if he invests in improvements of land or in buildings on it, he will not appreciably curtail the opportunities of others to invest capital in like improvements."

The underlying reason why land is not capital is that "land is a permanent and fixed stock while appliances made by man . . . are a flow capable of being increased or diminished . . . Now if the nation *as a whole* finds its stock of planing machines or ploughs inappropriately large or inappropriately small, it can redistribute its resources. It can obtain more of that in which it is deficient, while gradually lessening its stock of such things as are superabundant: *but it cannot do that in regard to land*: it can cultivate its land more intensively but it cannot get any more." (Emphasis in the original.) In other words, capital is a produced means of production; environment is not.[21]

Marshall's remarkable conclusion is that "these considerations lead us to repeat that . . . a far-seeing statesman will feel a greater responsibility to future generations when legislating as to land than as to other forms of wealth; and that from the economic and from the ethical point of view, land must everywhere and always be classed as a thing by itself."

Potvin has begun to consider the implications of equating the environment with capital in the ongoing attempts to extend existing macro-economic models (Potvin 1989) if environment is

---

[19] In this theory, the ownership of the factors of production – land, labour, and capital – is taken as given. When the market is in equilibrium, the owners of the factors of production receive a share of the total output of the economy that is determined by the value of the marginal product of the factors that they own. The owners of land, capital, and labour are only distinguished by the factors of production that they own; the principles upon which their share of output is based are identical.

[20] Marshall's comments on the distinction between land and capital are so interesting and well expressed that much of this section consists of direct quotations from his *Principles*.

[21] Arguably, renewable resources such as forestry and fish more closely resemble the neoclassical concept of capital. Although these resources can be depleted through excess harvesting, and the fertility of land itself can also be depleted, there is a virtual consensus in the literature that sustainable development calls for management objectives for renewable resources based on sustainable yields.

then valued in monetary terms and treated in the same manner as manufactured capital. Potvin points out some adverse implications for policy recommendations arising from such models if methods are used for inappropriately valuing the environment: the environment may be depleted and degraded on the mistaken assumption that an addition of manufactured capital of equal or greater monetary value will be forthcoming. His comments apply with equal force to similar attempts to incorporate a value for the environment in a modified system of national accounts.

If there are sound reasons for not treating the environment as just another form of capital, then alternative approaches must be sought to integrate the environment into economic analysis.

6 The Thermodynamic School – Economics and the Laws of Nature

Having suggested in previous sections that the problem of measuring the stock of natural capital requires further work, and that the treatment of the environment as capital is problematic, it is interesting to consider the efforts of economists such as Boulding (1966), Ayres and Kneese (1969), Victor (1972), Georgescu-Roegen (1979; 1975), Daly (1977), and Perrings (1987). Their contributions are referred to here as the thermodynamic school. These and other writers have attempted to show the implications of the basic laws of physics for understanding economics.

There are two streams of analysis to be considered. The first, which is sometimes known as the materials-energy balance approach, is based on the first law of thermodynamics – the law of conservation of matter-energy. According to this law, economic activity cannot create or destroy matter or energy, it can only rearrange it.[22] Consequently, all of the materials and energy used by the economy must ultimately be disposed of back into the environment. Their form may have changed, but the quantity of inputs from the environment must be balanced, at some time or another, by an equal quantity of residuals returned to the environment.

This perspective on the economy raises some interesting issues. It assumes some sort of clear demarcation between the economy and the environment, which in practice, may be difficult to define. For example, is a mineral in the ground part of the economy or does it only enter the economy when it is extracted for processing? How should buildings and roads be considered – as part of the economy or part of the environment? These are taxonomic questions highlighted by the materials-energy balance approach, but which need to be addressed in any attempt to situate the economy within its environmental context.

While most economists of this school have focused primarily or exclusively on the first law of thermodynamics, Georgescu-Roegen has argued that it is the second law of thermodynamics which is the most important for economics. He offers the following statement of the second law, for the nonspecialist: "Heat flows by itself only from the hotter to the colder body, never in reverse . . . It is thermodynamics . . . that recognizes the qualitative distinction which economists should have made from the outset between the inputs of valuable resources (low entropy) and the final outputs of valueless waste (high entropy)" (Georgescu-Roegen 1975: 351-3).

Thus, according to Georgescu-Roegen, and reinforced by Daly and Perrings, economic activity takes low entropy matter-energy (which can be thought of as highly organized materials and energy flows) and converts it into high entropy matter-energy (highly disorganized materials and energy). For example, the combustion of fossil fuels converts highly structured materials into widely dispersed gases and particles to gain access to the chemical energy bound up in the gas, oil, or coal. This energy is liberated as heat in an *irreversible* process (without an infusion of energy that is greater than the amount of energy liberated in the first place).

The entropic nature of economic activity explains the reason why 100% recycling within a closed system is an impossibility.[23] Taking the longest of long views, the second law of thermodynamics, from which the inevitability of increasing entropy derives, says that sustainable development is unachievable, even meaningless. But that is a

[22] Furthermore, except for the minuscule amounts of matter transformed into energy in nuclear reactors, the quantities of matter and energy counted separately are unaffected by economic activity.

[23] The Earth is effectively a closed system for matter but an open system for energy, receiving an enormous influx of energy from the sun.

very long view indeed, requiring a planning horizon of astronomical magnitude. Limiting ourselves to the resources of just one planet that is open to a continual and long-lasting influx of solar radiation, the direct relevance of the entropic perspective becomes questionable. It does, however, provide several valuable insights of more immediate significance.

First, it reminds us that there may be physical measures of natural capital that can be used to aggregate the capital stock independently of prices. The effort in the 1970s to develop procedures for energy accounting in which the direct and indirect energy used in all products can be estimated is but one example of such an approach.[24]

Second, as already noted, complete recycling of manufactured products is not possible. The economic system will always require a continuing input of resources from the environment and it will always generate residuals to be disposed of within the environment. In the context of sustainable development, where a time horizon of 100 years or more is not out of place, it becomes true that virtually all the manufactured capital stock of an economy must be replaced. The demands placed on the natural capital stock by such an endeavour are by no means trivial.

Third, in a growing economy, even if recycling of mineral ores could be increased from zero (for most materials except those used as fuels it is already far above zero) to 100% (an unobtainable objective), the reduced demands for additional ore from virgin sources would be equivalent to those required in one doubling period. This is illustrated in Table 1.

The example shows that in both cases the total use of resources is identical in each year. After 10 years, by which time the use of the resource has doubled, the use of the virgin resource is twice as high with no recycling as it is with 100% recycling. This relationship remains the same in later years. The effect of a shift from no recycling to complete recycling (which is considerably more than could possibly be achieved in real life), is to add one doubling period to the increasing demand for virgin resources. This is not inconsequential, but it does not save a diminishing resource stock from ultimate depletion.

Finally, the thermodynamic school reminds us to take the long view in our deliberations about sustainable development. What, for instance, is the significance of Georgescu-Roegen's observation that the industrialization of the past two centuries has been characterized by a shift in dependency on renewable resources with large stocks and flows, which are hard to regulate (e.g., solar energy), to a dependency on non-renewable resources with comparatively small stocks and flows that can temporarily be increased at will (e.g., fossil fuels)? Taking the long view is no easy task for economists since much of what can safely be held constant over shorter periods – the basic structure of the economy, the types of commodities produced, the legal and political structure within which the economy operates – will not remain unchanged in the future. Thus, it will be necessary for economists to have a firmer grasp on the

[24] This approach presents problems in dealing with qualitative differences in different forms of energy. Each energy product has more characteristics of value than simply joules.

Table 1: Hypothetical Example of the Effect of Materials Recycling on Resource Use in the Context of Growth (tonnes)

	Zero Recycling			100% Recycling		
Year	Use of Virgin Resource	Use of Recycled Resource	Total	Use of Virgin Resource	Use of Recycled Resource	Total
1990	10	0	10	10	0	10
2000	20	0	20	10	10	20
2010	40	0	40	20	20	40
2020	80	0	80	40	40	80

biophysical underpinnings of economic activity if they are to be credible participants in the discussion about sustainable development.[25]

6.1 Thermodynamic Indicators of Sustainable Development

Daly (1989) has suggested four operational principles of sustainability. He does not claim to have arrived at a completely worked-out set of principles, his aim being to contribute to their eventual development.[26] One aspect of operationality is measurability. Unless the effect of adopting and implementing each principle can be measured by an indicator of some sort, the operationality of the principle is questionable. It is useful, therefore, to review Daly's principles by considering their potential role as the basis for indicators of sustainable development.

1. *"The main principle is to limit the human scale to a level which, if not optimal, is at least within carrying capacity and therefore sustainable."*

   The measurement of carrying capacity lies beyond the purview of economics. If it can be measured, then according to this principle, it should not be exceeded by the combination of the number of people and the average level of per capita resource consumption. Both of these variables can be measured, but without the associated measurement of carrying capacity they have limited usefulness as indicators of sustainable development.

2. *"Technological progress for sustainable development should be efficiency-increasing rather than throughput-increasing. Limiting the scale of resource throughput would induce this technological shift."*

   If implementation of this principle requires that total throughput in an economy be measured, then it should be recognized that this would be a very onerous task. It makes more sense to concentrate on those inputs and outputs of materials and energy that link the economy and the environment and that are related to specific limiting factors.[27]

3. *For renewable resources, "harvesting rates should not exceed regeneration rates and waste emissions should not exceed the renewable assimilative capacity of the environment."*

   Harvesting rates and waste emissions are readily measurable and most economies already maintain partial records of these important variables. However, regeneration rates and the assimilative capacity of the environment, which are themselves not unrelated, are much more difficult to measure and existing data are far more sparse.

4. *"Non-renewable resources should be exploited, but at a rate equal to the creation of renewable substitutes."*

   This principle corresponds to Pearce and Turner's call to maintain the stock of natural capital and specifically addresses the problem of the inevitable depletion of non-renewable resources if they are used at all.[28] In the absence of renewable and non-renewable resources that are perfect substitutes, there are obvious difficulties in determining whether the creation of a renewable resource is adequate to offset the exploitation of a non-renewable resource. Daly offers no indication of how this might be decided. Nevertheless, the measurement of the use of non-renewable resources is reasonably straightforward, and far more so than the equally important task of measuring remaining resource stocks.

---

[25] Even Keynes, in his *General Theory of Employment, Interest and Money* (1936), confined himself to the short run with his celebrated comment that "in the long run we are all dead." Perhaps it is only now that we are beginning to understand what he really meant.

[26] An area which Daly has not addressed is the applicability of his principles for sustainability to specific regions of the globe. If they are to be applied at anything less than the global level, then an additional set of boundary conditions must be identified specifying and possibly limiting the extent to which development in any one region can depend on economic and environmental interactions with other regions.

[27] The implications of such limiting factors for economic activity in Ontario are the focus of attention in work under way in Ontario. See Appendix B.

[28] Implementation of this principle implies that the previous principle requires modification: harvesting rates should be maintained *below* regeneration rates to the extent necessary to compensate for the depletion of non-renewable resources.

7 Conclusion

This paper has approached the issue of indicators of sustainable development from the standpoint of capital theory. As such, it has benefited from some of the insights into the meaning of capital that have been obtained by economists over many years of careful deliberation. At the same time, it has suffered from the unresolved difficulties inherent in capital theory, the issues of substitution and aggregation being the main ones, which carry over into any attempt to apply capital-theoretic notions to sustainable development. Marshall's insights into the differences between land and capital were a lead-in to the ambitious work of the thermodynamic school, which stresses the importance of understanding the biophysical underpinnings of economic activity.

Acknowledgements

The author gratefully acknowledges comments received on an earlier draft of this paper from D. Butterfield, T. Burrell, E. Hanna, A. Kubersi, J. Potvin, and his other colleagues at CEAC. Additional revisions were made to the paper after the workshop.

References

Ahmad, S. Unpublished manuscript.

Ayres, R., and A.V. Kneese. 1969. Production, Consumption, and Externalities. *American Economic Review* 59:282-97.

Barnett, H.J., and C. Morse. 1963. *Scarcity and Growth: The Economics of Natural Resource Scarcity*. Baltimore and London: Johns Hopkins University Press.

Bartelmus, P. 1989. *Sustainable Development: A Conceptual Framework*. Department of International Economic and Social Affairs, Working Paper Number 13. New York: United Nations.

Boulding, K. 1966. The Economics of the Coming Spaceship Earth. In *Environmental Quality in a Growing Economy*, by H. Jarrett (ed.). Baltimore and London: Johns Hopkins University Press.

Brown, G.M., and B. Field. 1979. The Adequacy of Measures for Signaling the Scarcity of Natural Resources. In *Scarcity and Growth Reconsidered*, by V.K. Smith (ed.). Baltimore and London: Johns Hopkins University Press.

Burmeister, E. 1980. *Capital Theory and Dynamics*. London: Cambridge University Press.

Daly, H. 1977. *Steady-State Economics*. San Francisco: W.H. Freeman and Company.

Daly, H. 1989. Sustainable Development: From Concept and Theory Towards Operational Principles. *Population and Development Review*. Special Issue on the 1989 Hoover Institution Conference.

Dasgupta, P., and D. Heal. 1979. *Economic Theory and Exhaustible Resources*. London: Cambridge University Press.

Fisher, A. 1981. *Resource and Environmental Economics*. London: Cambridge University Press.

Georgescu-Roegen, N. 1975. Energy and Economic Myths. *Southern Economic Journal*, 41:3.

Georgescu-Roegen, N. 1979. Comments on the Papers by Daly and Stiglitz. In *Scarcity and Growth Reconsidered*, by V.K. Smith (ed.). Baltimore and London: Johns Hopkins University Press.

Harcourt, G.C. 1972. *Some Cambridge Controversies in the Theory of Capital*. Cambridge: Cambridge University Press.

Hamouda, O.F., and G. Harcourt 1988. Post-Keynesianism: From Criticism to Coherence? *Bulletin of Economic Research* 40:1.

Herfindahl, O., and A.V. Kneese 1974. *Natural Theory of Natural Resources*. Columbus, Ohio: Charles E. Merrill Publishing Company.

Hotelling, H. 1931. The Economics of Exhaustible Resources. *The Journal of Political Economy*, vol. 39.

Keynes, J. M. 1936. *The General Theory of Employment, Interest and Money*. Cambridge: Cambridge University Press.

Krutilla, J.V. 1967. Conservation Reconsidered. *American Economic Review* 47:777-86.

Marshall, A. 1920. *Principles of Economics*, 8th Edition. London: Macmillan.

Pearce, D., and R. Turner 1990. *Economics of Natural Resources and the Environment*. New York: Harvester Wheatsheaf.

Perrings, C. 1987. *Economy and Environment: A Theoretical Essay on the Interdependence of Economic and Environmental Systems*. Cambridge: Cambridge University Press.

Pezzy, J. 1989. *Economic Analysis of Sustainable Growth and Sustainable Development*. The World Bank, Environment Department Working Paper Number 15.

Potvin, J.R. 1989. Alternative Notions of Geosphere/Biosphere in Economics: A Note on Words and Meanings. Hull, Quebec: Environment Canada. Photocopy.

Potvin, J.R. 1990. *Economic-Environmental Accounts: A Conspectus on Current Developments*. Ottawa: Rawson Academy of Aquatic Science, Occasional Papers Series, no. 3.

Scott, A.D. 1955. *Natural Resources: The Economics of Conservation*. Toronto: University of Toronto Press.

Slade, M.E. 1982. Trends in Natural Resource Commodity Prices: An Analysis of the Time Domain. *Journal of Environmental Economics*, vol. 9, no. 2.

Smith, V.K. (ed.). 1979. *Scarcity and Growth Reconsidered*. Baltimore and London: Johns Hopkins University Press.

Solow, R. 1974. Richard T. Ely Lecture: The Economics of Resources or the Resources of Economics. *American Economic Review* 64:1-14.

Stiglitz, J.E. 1979. The Viability Issue: Is Sustainable Growth Possible in the Presence of Exhaustible Natural Resources? In *Scarcity and Growth Reconsidered*, by V.K. Smith (ed.). Baltimore and London: Johns Hopkins University Press.

Victor, P. 1972. *Pollution: Economy and Environment*. Toronto: University of Toronto Press.

Appendix A: Neoclassical Theory – A Closer Look

Section 2 of this paper outlined some of the main ideas from neoclassical economics that are relevant to sustainable development. This appendix gives a more detailed account of the same material, drawing heavily on the work of Dasgupta and Heal (1979).

Dasgupta and Heal's analytical model is as follows. They use a Cobb-Douglas production function to describe the production technology of the economy at the aggregate level:[29]

1. $Y = K^aR^bL^c$ with $a,b,c > 0$ and $a + b + c = 1$

where: Y = output
K = capital
R = resources
L = labour

One of the features of a Cobb-Douglas production function is that the elasticity of substitution between the inputs is equal to one. This elasticity gives the percentage change in the ratio of capital to resources that would be brought about by a percentage change in their relative prices. It is the concept most often used by economists to measure the ease with which inputs can be substituted. In a Cobb-Douglas production function, an x% increase in the ratio of the prices of resources and capital would lead to an x% decrease in the ratio of the quantities of resources and capital.

Returning to the question of sustained production with exhaustible resources, Dasgupta and Heal show that, with a Cobb-Douglas production function, "it transpires that the crucial question is whether or not a > b. This should not be entirely surprising. For these two parameters represent the elasticities of output with respect to reproducible capital and the exhaustible resource. Roughly speaking, if a > b, fixed capital is sufficiently important in production to allow for the possibility of a permanently maintainable output level despite the declining availability of the natural resource. The idea is to accumulate capital at a sufficiently fast rate to make this feasible" (Dasgupta and Heal 1979: 200).

Although Dasgupta and Heal's book is explicitly concerned with theory, rather than empirical issues, they do venture the opinion that 'a' is indeed larger than 'b' "which is presumably the best educated guess today. From the evidence of factor shares alone (which in a Cobb-Douglas world with perfect competition are equal to the relevant exponents in the production, one would imagine a to be about four times b" (1979: 205). Thus, *even without technological progress*, a depleting stock of resources does not appear to present a problem for sustainable development and economic output can be maintained or even increased indefinitely. When the promise of technological progress is introduced, the possibilities for sustained increases in output improve even more.[30]

A.1 Some Difficulties with the Neoclassical Approach

It should be obvious that since resources are required to manufacture capital goods, then the success of any attempt to substitute manufactured capital for resources will be limited by the extent to which the increase in manufactured capital requires an input of resources. The situation may be exacerbated if the operation of the capital requires further inputs of resources either to provide energy or materials with which to work, or both.[31]

In light of the significance given to the results obtained by Dasgupta and Heal, and other prominent economists as well, it is worthwhile to see how these results are affected when the capital and resources are distinguished by recognizing that resources are a gift of nature and capital requires resources in its production. This can be accomplished by defining a second equation to accompany equation 1 presented earlier:

2. $K = K^dR^eL^f$ with $d,e,f > 0$ and $d + e + f = 1$

Equation 2 is a Cobb-Douglas production function showing how capital is produced from capital,

---

[29] Dasgupta and Heal exclude labour from their production function since, by assumption, the input of labour is held constant. In this appendix, labour is included in the production function for completeness, though it does not affect any of the results.

[30] If the elasticity of substitution exceeds unity, *over the full range of possible combinations of capital and resources*, then a constant level of per capita consumption can be sustained even if the resource base is declining, and even if the share of income received by owners of resources 'b' exceeds that of the owners of capital 'a'.

[31] Several writers, such as Daly (1989) and Bartelmus (1989) have emphasized the difference between capital that can substitute for resources and capital that is complementary to resources.

resources and labour. Substituting equation 2 into equation 1 gives equation 3:

3. $Y = K^{ad}R^{(ae+b)}L^{(af+c)}$

Equation 3 is a new aggregate production function that is also Cobb-Douglas in form, but which incorporates one of the two distinguishing features of capital: it is a produced means of production.[32] Following Dasgupta and Heal, the potential for capital to substitute for resources so that output is maintained indefinitely, despite the declining stock of resources, depends on whether $ad > ae + b$, or equivalently, whether $a(d - e) > b$.

By assumption, d and e are both positive and less than 1. Therefore, $d - e < 1$. It follows, therefore, that $a > b$ does not imply $a(d - e) > b$. Hence, even though the values of a and b may satisfy Dasgupta and Heal's condition for substitution as a solution to resource depletion, they may not satisfy the equivalent condition required when capital is produced from resources. Obviously, the situation becomes even more bleak when the requirements for resources to operate capital are introduced.

This analysis can be taken a step further. By solving equation 2 for capital and substituting the expression for capital into equation 1, it is possible to obtain another Cobb-Douglas production function in which output is a function only of resources and labour:

4. $Y = R^{(ae/(1-d)+b)}L^{(af/(1-d)+c)}$

In equation 4, capital has disappeared from the equation altogether and output is related exclusively to the non-produced resources. Any possibility for solving a problem of depleting resources through substitution with capital has evaporated.[33]

A second limitation of the results from the neoclassical analysis of Dasgupta and Heal, of which they and other economists are fully aware, is that the potential for substitution of capital for resources may decline with increases in the ratio of capital to resources. Most neoclassical formulations of the production function, including the Cobb-Douglas function, assume that this will not be the case. Likewise, it is conceivable that technological progress may be subject to diminishing returns and will eventually become incapable of compensating fully for continuing depletion of the resource stock. Some economists argue that this situation exists particularly for the capacity of technological progress to provide substitutes for some services of the natural environment.[34]

A third limitation of this type of analysis has been stressed by Georgescu-Roegen (1979), who argues that production functions of the type shown as equations 1 and 2 confuse 'flow' variables with 'fund' variables. For example, the input of labour 'L' is best thought of in terms of person-days, which is a measure of the flow of services provided by the work force. Similarly, R is a measure of the flow of services from natural resources. K, however, is a measure of the stock of capital and not its services. Hence, L, R, and K are not directly comparable.[35]

---

[32] This formulation lacks a time dimension, which should be included in a more complete treatment.

[33] I am grateful to Dave Butterfield for this result.

[34] Krutilla and others have made much of the asymmetrical role played by technical progress. Although it can enhance the capital stock of the economy, and thereby increase the economy's capacity to produce commodities, it is far less capable of enhancing the supply of wilderness (Krutilla 1967).

[35] "The slipshod manner in which the factors are defined is another consequence of the infatuation with paper and pencil exercises" (Georgescu-Roegen 1979: 98).

Appendix B: Indicators, Models, and Constraints – An Outline of Work in Progress

A great deal of work is under way all over the world on defining and measuring sustainable development. International agencies such as the United Nations and the World Bank are developing new approaches to the system of national accounts so that they more appropriately deal with environmental degradation and resource depletion. Similar work is proceeding at the national level, including Canada, where interest in these matters is also being shown at the provincial and territorial levels. One effort, which is briefly described in this appendix, is a project recently initiated in Ontario by a team of economists and environmental scientists of which the author of this paper is a member.[36]

There are several distinguishing features of this approach including the way in which capital, in all its forms, is dealt with:

1. Underlying the economic production system there is a natural production system that imposes limits to the economic system. These limits are represented by a set of six fundamental environmental factors, which provide the ultimate constraints on economic activity and which must be respected in any program of sustainable development: surface area, air, water, minerals, the gene pool, solar radiation. Ultimately, it is the efficiency with which these factors are used that limit what can be achieved through sustainable development.

2. A vast amount of data on the Ontario economy, including the demands it makes on the above six factors and resources at risk (e.g., humans, flora and fauna, buildings and structures) is being assembled into a relational database.

3. An accounting framework is being established that incorporates economic and non-economic data within a coherent and systematic framework.

4. An enhanced input-output model of the Ontario economy is being built, which incorporates the accumulation of manufactured capital, allows for different technologies to be used by various economic sectors, locates economic activity at the sub-provincial level, explicitly includes the environmental protection and recycling industries, and incorporates quantitative linkages between the sectors of the economy and the six fundamental environmental factors.

Despite the rich literature that exists on the meaning of sustainable development, there is no agreement as to what would constitute a satisfactory set of indicators of sustainable development. However, it is generally acknowledged that one of the main features of sustainable development is that environmental constraints on economic activity should be respected. The lack of consensus on indicators implies a lack of consensus on the nature and specifics of these constraints.

Much of this paper has been concerned with the idea that, in some sense, maintenance of the stock of capital is a condition of sustainable development. Hence, one avenue for delineating indicators of sustainable development is to explore the meaning of a constant capital stock, particularly when possibilities for substitution among its constituent parts are allowed. Various insights into this matter have been gleaned from an examination of different schools of thought in economics but, unfortunately, many questions remain to be answered before a theoretically sound and implementable approach to measuring the capital stock will emerge.

B.1 A Different Approach to Indicators

The approach to sustainable development indicators described here is indirect. It allows for an analysis of the implications of different sets of constraints without making a prior commitment to any one set. An objective function will be established along somewhat traditional lines, for example, maximization of value added. Achievement of this objective will be constrained by the aggregate production functions of the economy (as expressed by the standard input-output relations with some allowance for alternative technologies). Additional constraints will be imposed to limit the extent to which the economy can impinge on each of the six fundamental environmental factors and any other factor believed to limit economic activity. Hence,

[36] This work is being undertaken by VHB Research and Consulting Inc. and Econometrics Research Limited (VHB-ERL) under contract to the Ontario Ministry of the Environment. The ministry is not necessarily in agreement with the views presented here.

there may be constraints on the amount of labour available, the size of the trade deficit, as well as the permitted levels of effluents disposed of in the environment and the resources obtained therefrom.

When the model is solved, a shadow price will be found for each constraint. If a constraint is not binding, in the sense that it does not limit the achievement of the objective, its shadow price is zero. If a constraint is binding, its non-zero shadow price will show how a unit change in the level of the constraint would affect the level of value added in the economy. Overall, the imposition of constraints on the model, which are intended to capture the meaning and intent of sustainable development, will help show the scale of adjustment that may be required to transform the existing economy into one that is sustainable. One measure of this adjustment will be the difference in total and sectoral value added between what can be achieved with and without the environmental constraints.

Taking a modelling approach to sustainable development has other advantages. For example, it will be possible to take the set of shadow prices and impose the requirement that the prices of all commodities in the economy fully cover not only the costs of production normally recognized in market exchanges, but also the demands made on the environment at all stages of production, consumption, and disposal. The model will generate a new set of 'full-cost' prices, and a comparison of these prices with those prevailing in the marketplace will provide another indication of the degree of adjustment required to achieve sustainable development. Further, including a set of price-dependent demand equations will make it possible to reiterate the calculations until an equilibrium set of full-cost prices and quantities of all types is identified that is consistent with the initial set of constraints. Conceivably, these full-cost prices could be used in cost-benefit analysis with some assurance that the outcome will be consistent with the constrained requirements of sustainable development.

The Concept of Ecological Integrity, Alternative Theories of Ecology, and Implications for Decision-Support Indicators

James J. Kay

Abstract

The subject of this paper is indicators of ecological sustainability. To discuss this subject, an appreciation of the state of ecology is required. Ecology as a science is at a formative stage, in much the same state as physics at the time of Galileo. We await the insight of a Newton. There is little consensus in theoretical ecology about what the important characteristics of ecosystems are (McIntosh 1985). In spite of this, there is the beginning of an empirical understanding of the effects of environmental change on ecosystem development and hence sustainability.

This paper will explore some recent advances in our understanding of the development of complex systems, their implications for our understanding of how ecosystems develop, a framework for discussing ecosystem integrity, and current theories related to integrity and the measures they suggest. A basic premise of this discussion is that if economic development is to be sustainable then it should enhance the integrity of ecosystems, or at the very least not erode ecosystem integrity.

Contents

1 The Development of Complex Systems....24

2 A Framework for Describing Responses to Environmental Change....26

3 Different Ecological Theories and Their State Variables....31

4 Some Examples of Indicators and Their Use....37

5 Concluding Remarks....46

References....47

Appendix A: Some Systems Notions....50

Appendix B: Ecological Stability....55

KAY, J. J. 1991. The Concept of Ecological Integrity, Alternative Theories of Ecology, and Implications for Decision-Support Indicators. In: CANADIAN ENVIRONMENTAL ADVISORY COUNCIL. *Economic, Ecological, and Decision Theories: Indicators of Ecologically Sustainable Development.* Ottawa: CEAC. Prepared under contract for the CEAC Workshop on Indicators for Ecologically Sustainable Development (Ottawa, July 1990).

1 The Development of Complex Systems

To begin, let us consider how non-equilibrium thermodynamics suggests that systems develop. Prigogine has shown that under certain conditions, open systems with a gradient across their boundaries will move away from equilibrium and will establish new stable structures (Prigogine and others 1972, Nicolis and Prigogine, 1977, 1989). This is the opposite of the behaviour one would normally expect given the second law of thermodynamics[1]. Such systems are characterized by rates of energy dissipation, which increase as the system moves from equilibrium and becomes more organized. Hence the name *dissipative structures*. Simple examples are vortices in bathtubs, tornados, and lasers.

The development of such self-organizing systems is characterized by phases of rapid organization to a steady-state level followed by a period during which the system maintains itself at the new steady state. The organization of the system is not a smooth process but rather proceeds in spurts. These spurts are a sudden acceleration in the change in the state of the system. The state change may be continuous or catastrophic[2]. The change in the state is accomplished by the addition of new dissipative structures to the system. These new structures can consist of new pathways for energy flow that connect old components, or new components and their associated new pathways. Each spurt results in the system moving further from equilibrium, dissipating more energy, and becoming more organized. Each spurt occurs when random environmental conditions exceed a catastrophe threshold for the system. The path through state space which the system follows as it develops in a stable environment, is called the 'thermodynamic branch'. Ecosystem succession is an example of this kind of process. Each of the serial stages corresponds to one steady-state plateau. The displacement of a previous serial stage by the next is an example of a spurt, the reorganization of the system to a new level of structure that dissipates more energy.

The gradient that drives ecosystem development is the solar energy impinging on the ecosystem (Kay 1984, 1989). As ecosystems are driven away from equilibrium they become more organized[3] and effective at dissipating solar energy. At the same time as this self-organizing process is occurring in ecosystems, external environmental[4] fluctuations are tending to disorganize the system. The multidimensional state space for which the disorganizing forces of external environmental change and the organizing thermodynamic forces are balanced is referred to here as the 'optimum operating point' (see Figure 1).

For any real ecosystem, a particular point will be an optimum operating point only temporarily. This is because the external environment will be in flux and evolution will be proceeding, thus changing the balance between the organizing and disorganizing forces. However it is useful over short time periods to treat the optimum operating point as if it were stationary. The climax community in ecological succession would be an example of an optimum operating point for an ecosystem. The climax community represents a temporary balance between the organizing forces in ecosystems and the disorganizing forces. Over evolutionary time, however, new species will enter the equation – so will new environmental phenomena. This will result in a new climax community, and a different optimum operating point.

In the context of these ideas, our sense of the system as a whole, that is, its integrity, has to do with the ability of the system to maintain its organization and to continue its process of self-organization. If a system is able to maintain its organization in the face of changing environmental conditions then it is said to have integrity. If a system is unable to maintain its organization then it has lost its integrity. Integrity is the ability of the system to attain and maintain its optimum operating point.

This is an important implicit aspect of the definition of integrity of which the reader must be aware. Ecosystems are not static. Their organization is often changing, both in the short-

[1] The Second Law of Thermodynamics: Entropy will increase; the state of a system will tend to disorder; systems tend to equilibrium.

[2] Appendix A discusses a set of system notions such as state, stability, catastrophe, and equilibrium.

[3] Change in organization refers to changes in the function of a system and its internal connections (structure) so as to better carry out some organizational imperative.

[4] Environment is used throughout the paper in its system theoretic sense, that is, those things that are not part of the system but that have an effect on it. Thus environment refers to the biotic and abiotic components external to an ecosystem that affect it. Humans may be considered as part of the environment of an ecosystem or as part of the ecosystem itself, depending on the context.

term and in an evolutionary sense. Furthermore, any loss of organization is often gradual. Thus it is not possible to identify a single organizational state of the system that corresponds to integrity. Instead there is a range of organizational states for which the ecosystem is considered to have integrity. For the sake of discussion, the optimum operating point is treated as a single stationary point in multidimensional state space. In reality, it is a set of points in state space whose membership changes over time. The definition of this set would necessarily have a human component.

Also, the reader should be aware that the theory of dissipative structures suggests that a number of different organizational and developmental pathways are available to ecosystems. These pathways are non-linear and may be discontinuous and multi-valued. Thus the (set of) optimum operating point(s) or set point for the system is not unique. There will be several different possible (sets of) optimum operating point states for the system. As Holling (1986) has shown, the normal course of ecosystem development can consist of flips in state, that is catastrophic changes. Detailed discussion of these notions can be found in Appendix A.

What follows is a discussion of the use of these notions to examine ecosystem response to environmental change and the implications for integrity as a concept central to sustainable development indicators.

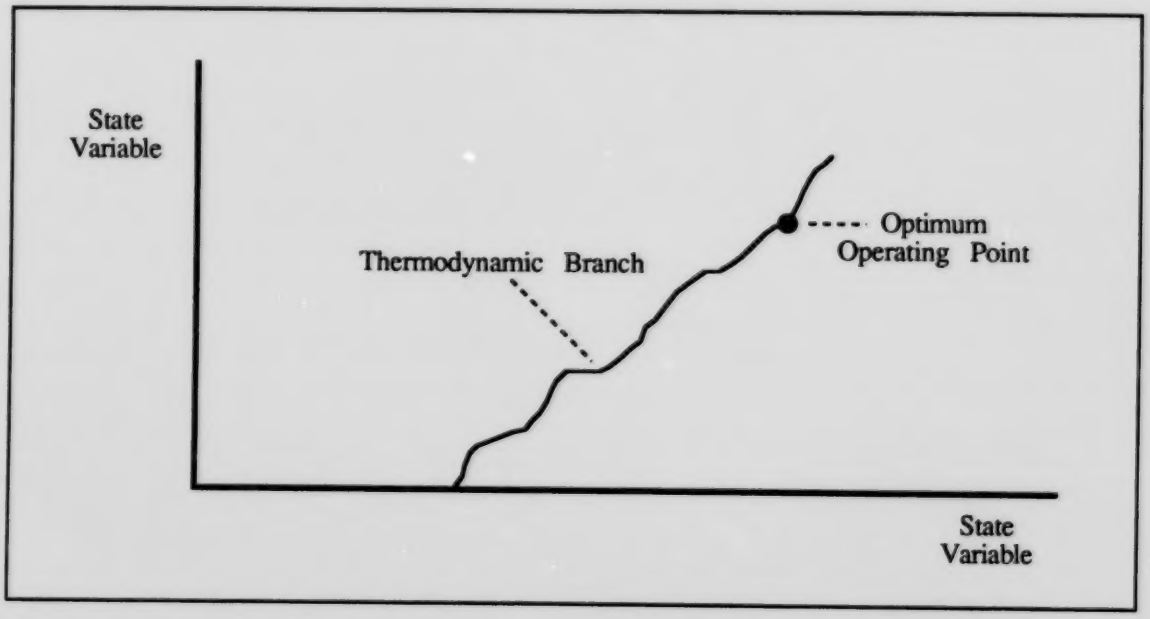


Figure 1: An ecosystem develops along a thermodynamic branch (a path in two-dimensional state space) until it reaches an optimum operating point. Examples of state variables could be net productivity and biomass.

2 A Framework for Describing Responses to Environmental Change[5]

Let us assume that the ecosystem has developed along a thermodynamic branch in the way described above and that it has reached its optimum operating point. Suppose some change occurs in its environment. (The change may be short-term with the environment returning to its previous condition, or the change may persist.) What effect will this have on the ecosystem's organization and hence its integrity? A series of questions must be asked:

Will the system move away from its optimum operating point?

- If the response is **no**, then organization and integrity are not immediately affected.
- If the response is **yes**, then the question becomes:

Does the system return to its original optimum operating point?

- If the answer is **yes**, then there are three issues:

1. How far did the system move from its optimum operating point before returning?
2. How long did it take to return to its optimum operating point?
3. What is the stability of the system upon its return?

In any case the system reorganizes itself to cope with the environmental change and its integrity, for the moment, is preserved.

- If the answer is **no**, then there are two possibilities: a new optimum operating point exists or it does not. If there is no new optimum operating point, then the organization breaks down and the system loses its integrity (Case 0). If there is a new optimum operating point, then there are three possibilities:

Case 1 – the new optimum operating point is on the original thermodynamic branch;

Case 2 – the new optimum operating point is on a bifurcation from the original branch; or

Case 3 – the new optimum operating point is on a different thermodynamic branch and the system undergoes a catastrophic reorganization to reach it.

2.1 Examples and Elaboration

The ecosystem does not move from its original optimum operating point.

For example, consider a terrestrial ecosystem adapted to a periodic flood or drought. If the disturbance is not particularly intense the system will not be affected. Another example is the ongoing spraying of Fenitrothion on Canadian forests to control spruce budworm. This appears to have no immediate effect on the forests but it does interfere with the ability of the forests to regenerate (Weinberger and others 1981). Thus the ability of the system to deal with some future stress requiring regeneration may be impaired. Consequently this human activity may be unacceptable as it could ultimately affect the ecosystem's integrity.

The ecosystem moves from its original optimum operating point but returns to it.

Fire in a temperate, forested ecosystem is a short-term event that moves the system well away from its optimum operating point. However, the forest regenerates to the original system. Oil spills along the shores of Great Britain have had similar effect with a regeneration time of about 10 years (Nelson-Smith 1975). Rutledge (1974) showed that a shortgrass prairie ecosystem subjected to continuous drought reorganizes itself so that, after about 20 years, it will return to its pre-drought state.

The system moves permanently from its original optimum operating point.

Case 0: The system collapses.

The environment changes and becomes uninhabitable. An example is the process of desertification. Another is severe, prolonged drought in mangrove systems that leads to the total collapse of the system (Lugo and others 1981). A third example is the deposit of

[5] Most of this section, including the figures, is verbatim from: Kay, J., A Non-Equilibrium Thermodynamic Framework for Discussing Ecosystem Integrity, *Environmental Management*, in press.

substances by precipitation that, in the extreme case of the Sudbury area in Canada, has led to the rocky equivalent of a desert. In the Laurentian Shield, this has led to dead lakes. In each of these cases some life forms may continue to persist, but a complex ecosystem does not.

Case 1: The system remains on its original thermodynamic branch.

The ecosystem maintains its original set of dissipative structures (e.g., species), or moves back to some set that represents an earlier stage in development (see Figure 2). The level of operation of the individual structures has changed, perhaps even catastrophically. Overall the dissipative system is recognizable as the original, but its operation has been modified.

In this case, there are four issues:

1. How far is the new optimum operating point from the old?
2. How long does it take to reach the new optimum operating point?
3. What is the stability of the system about the new optimum operating point?
4. If the environmental conditions return to their original state, will the system return to the original optimum operating point?

Although the system's organization has changed in this case, it will probably return to the original optimum operating point if the environmental conditions return to their previous state. This is because all the original structures (e.g., species) exist to some extent. The system's integrity has been affected in the sense that its organization has had to change. This is noteworthy only if the new optimum operating point (level of operation) is considered undesirable.

As an example, consider the practice of spraying the end product of the secondary treatment of municipal waste water on terrestrial ecosystems. Pine forests subjected to such spraying are shifted back to an old field community – the developmental stage prior to a forest (Shure and Hunt 1981). The ecosystem is still on the original thermodynamic branch, but at an earlier stage of development. As another example, consider maple forests subjected to acid rain. They are shifted to a state of less productivity and lower biomass[6]. A cold snap in 1962-63 when the shoreline systems in Southern England were driven back to an earlier stage of development is another example. Recovery to the original state seems unlikely (Nelson-Smith 1975).

Case 2: There is system bifurcation to a new thermodynamic path.

In this case some new dissipative structures are added to the system and/or some of the original ones disappear (see Figure 3). The new structures can be new pathways for energy flow connecting old components, or the emergence of new components and their attendant pathways. Also, the level of operation of the system is changed. The system is seen as slightly different from the original.

The same four questions apply here as apply to Case 1 (see Table 1). However the answer to the fourth question is probably different. The system is not likely to return to its original optimum operating point, unless the bifurcation point is the original optimum operating point. It is theoretically possible, however, to return to the original optimum operating point by manipulating the environmental conditions. If it is not likely to return to its original optimum point, then the organization of the system has probably been permanently altered by the addition of new dissipative structures. However, bifurcations represent variations on the original theme, so the new ecosystem's organization will not be extraordinarily different from the original. The integrity of the system has been affected in the sense that the organization has been permanently altered, although not dramatically. Again, this is noteworthy only if the bifurcation branch and the new optimum operating point are considered undesirable.

An example of this case is the change in a marsh gut ecosystem in Crystal River, Florida (Kay 1984, Ulanowicz 1986). The system is stressed by warm water effluent from a nuclear power station (6 °C increase in water temperature). The result is the loss of two top predators, two lower predators, the addition of a three lower predators and a herbivore species, and a dramatic change in the foodweb for cycling and trophic positions.

---

[6]Unfortunately the level of acidity in the rain is increasing with attendant further changes in the ecosystems. The question is whether the response of the maple forests will remain as in Case 1 or become one of the other cases discussed here. If it does not remain as in Case 1, it would imply the loss of some of the characteristics of these forests that we value.

Table 1: The Possible Responses of an Ecosystem to Environmental Change

a) The ecosystem does not move from its original optimum operating point.

b) The ecosystem moves from its original optimum operating point but returns to it.

Issues concerning integrity to be considered for this case:

1. How far did the system move from its optimum operating point before returning?
2. How long did it take to return to its optimum operating point?
3. What is the stability of the system upon its return?

c) The ecosystem moves permanently from its original optimum operating point.

Case 0: The ecosystem collapses.

Case 1: The ecosystem remains on its original thermodynamic branch (see Figure 2).

Case 2: There is system bifurcation to a new thermodynamic path (see Figure 3).

Case 3: The system moves to a new thermodynamic branch (see Figure 4).

Issues concerning integrity to be considered for each of these four cases:

1. How far is the new optimum operating point from the old?
2. How long does it take to reach the new optimum operating point?
3. What is the stability of the system around the new optimum operating point?
4. If the environmental conditions return to their original state, will the system return to the original optimum operating point?

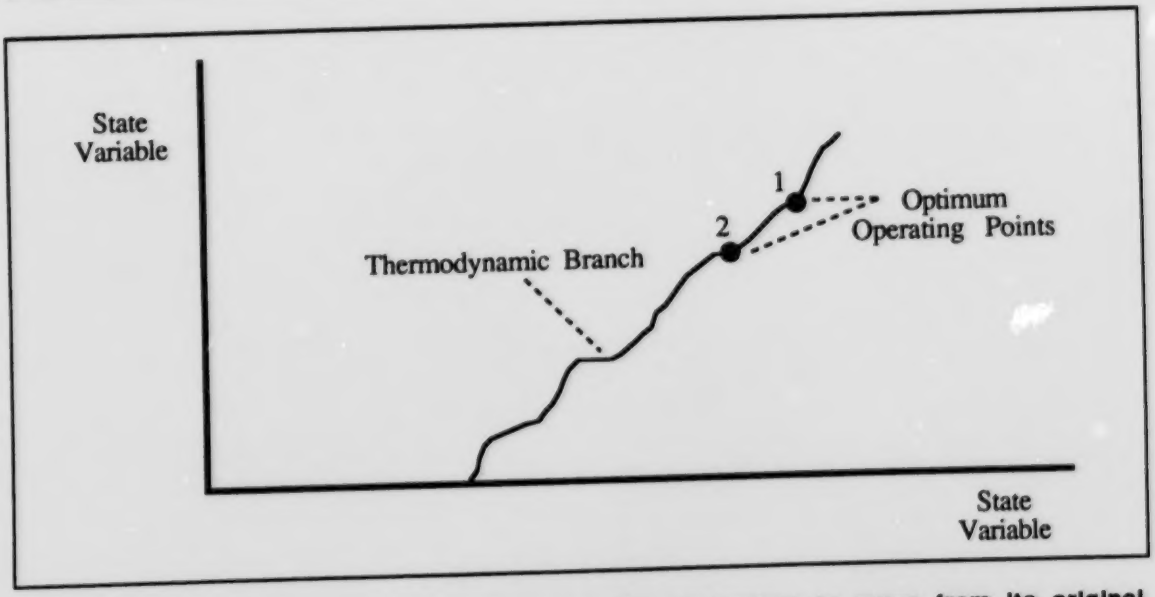


Figure 2: The environmental change causes the ecosystem to move from its original optimum operating point (1) to a new optimum operating point (2). An example of this would be a stress that causes an ecosystem to return to an earlier successional stage.

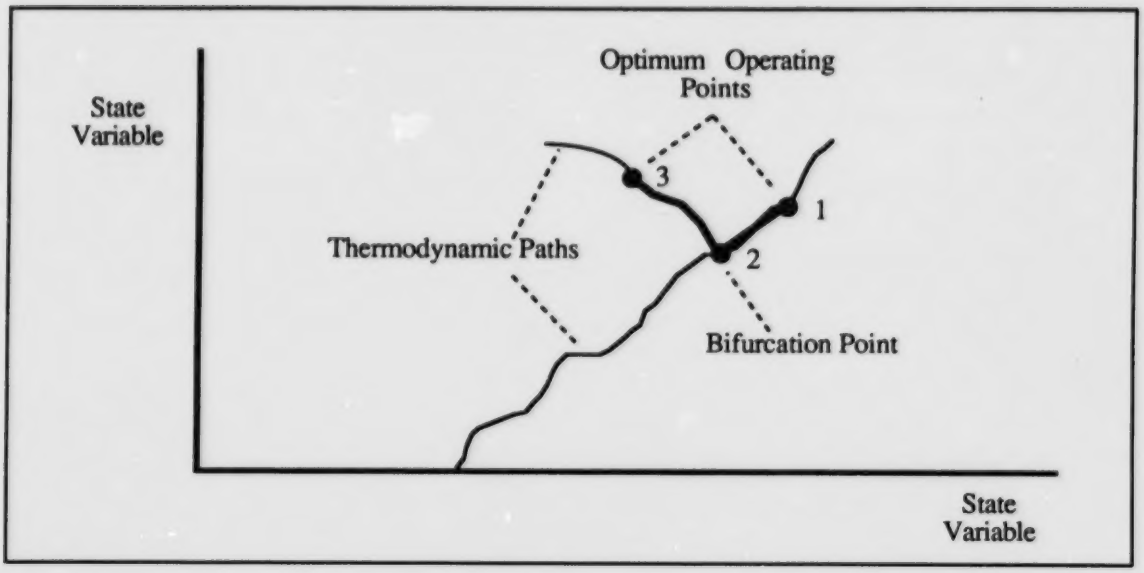


Figure 3: In response to changing environmental conditions, the system moves away from the original optimum operating point (1) through a bifurcation point (2), onto a new path, and then to a new optimum operating point (3). An example of this is a stress, such as the hot water effluent from power plants, which would result in a change in species composition.

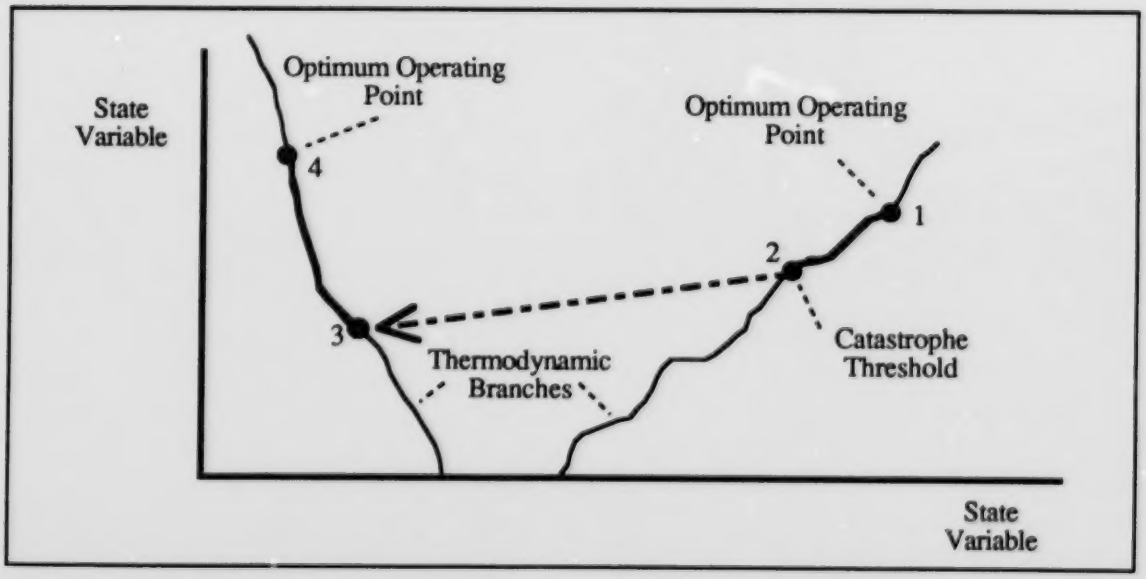


Figure 4: The environmental change drives the ecosystem from its original optimum operating point (1) through a catastrophe threshold (2) to a new thermodynamic branch at (3) and eventually to a new optimum operating point (4). An example is the elimination of fish in lakes because of acid rain.

These are examples of changes in the dissipative structures in an ecosystem. State variables such as net productivity decrease, thus the overall functioning of the system has changed. However, overall the ecosystem is clearly a variation of the original. It is not clear that stopping the effluent would result in a return to the original system. Similar results have been found for Par Pond on the Savannah River in South Carolina (Sharitz and Gibbons 1981).

Another example is the switch from a white spruce community to a black spruce community when there is a sharp reduction in nutrient availability. In these forested taiga ecosystems, black spruce are better suited to low nutrient situations and once established tend to exclude white spruce by maintaining the low nutrient situation. The white spruce is not able to reassert itself once displaced (DeAngelis and others 1989).

A final example is the introduction of exotics into the Great Lakes. New species associations (dissipative structures) occur. The sea lamprey is a case in point. It appears the system has been permanently altered, but it still resembles the original.

Case 3: The system moves to a new thermodynamic branch.

In this case, the system undergoes a catastrophic change that leaves the system so reorganized that it is clearly different from the original system (see Figure 4). There is no possibility of the system returning to its original optimum operating point, even if the environmental conditions return to their original state. (This is a hypothesis. In this case, the system is made up of very different dissipative structures than existed in the original. The author has been unable to find a single example of an ecosystem flipping back after undergoing such a dramatic reorganization.) In one sense, the integrity of the system has been seriously undermined, as the system will be quite different from the original. However, an ecosystem still exists, so in some sense, it has maintained its integrity.

Clearwater Lake is an excellent example of this phenomenon (Stokes 1984: 246-249). This lake is subjected to acid rain and has a pH of 4.3. Consequently there are no fish in the lake. The phytoplankton-zooplankton balance has shifted significantly, and "the system has flipped from one domain of stability to another." The ecosystem has reached a new optimum operating point on a new thermodynamic branch and further change is unlikely. Even if the acid rain stopped, the ecosystem would be unlikely to return to its former state without human intervention.

Another example of this case is the clear-cutting of a terrestrial system, such as a tropical rain forest. Soil erosion is so severe it effectively changes the soil type and precludes the original system from reappearing. Another example is a burn in a spruce hardwood forest on thin soils. This has resulted in a new bare rock-shrub ecosystem appearing as the climax (Bormann and Likens 1979). A final example is the irreversible change of savanna ecosystems to woody vegetation brought on by cattle grazing (Walker and others 1981).

2.2 Summary

In this section, the relationship between an ecosystem's integrity and its ability to maintain its organization has been explored from the perspective of dissipative structures. The possible organizational changes in response to environmental change were enumerated. The ways that such changes might be associated with changes in integrity of the ecosystem were examined. There are four points to note.

First, *dissipative systems can respond to environmental change in qualitatively different ways.* One response is for the system to continue to operate as before, even though its operations may be initially and temporally unsettled. A second response is for the system to operate at a different level using the same dissipative structures it originally had (for example, a reduction or increase in species numbers). A third response is for some new structures to emerge in the system to replace or augment existing structures (for example, new species or paths in the food web). A fourth response is for a new dissipative system, made up of different structures, to emerge. We must be aware of each of these possible responses, if we are to anticipate the impact of environmental change on the integrity of an ecosystem.

Second, if the concept of integrity is to be useful, it must have an anthropocentric component that indicates which changes in the ecosystem are considered acceptable by the human participants. The framework indicates ways in which an ecosystem might reorganize in the face of environmental change, but not which reorganizations constitute a loss of integrity.

It could be argued (and often is) that any environmental change that permanently changes the optimum operating point affects the integrity of the ecosystem. In this case, there would be four distinct types of loss of integrity – cases 0 through 3. It could also be argued that any time that the system can maintain itself at an optimum operating point, it has integrity. In this case, loss of integrity would occur only if the system is unable to maintain itself at an optimum operating point.

In between these two extreme positions there is the possibility of defining some optimum operating points as being undesirable changes in the system, and therefore representing a loss of integrity. This would inject an anthropocentric component into the definition of integrity. Without this anthropocentric component, we are restricted to defining integrity as the ability of an ecosystem to absorb environmental change without any ecosystem change. This would rule out the acceptability of the other three ecosystem responses to environmental change discussed above. This does not seem reasonable to this author. Ultimately an evaluation of the ecological sustainability of an action will depend on a value judgement concerning which changes in the affected ecosystem are acceptable to the human participants.

Third, and this follows from the second point, from an anthropocentric perspective, a new optimum operating point on a new thermodynamic branch may be preferred over the old optimum operating point on the system's previous branch. This may be the case even if it implies the catastrophic collapse of the previous ecosystem. The implication is that under some circumstances we might be willing to accept, even require, some degradation in existing ecosystems to allow a new one to flourish. The difficulty is that, by their very nature, bifurcations and flips can be unpredictable. When an ecosystem has been seriously degraded by an environmental change, returning it to its original state may not be the sensible management goal. Rather, aiming the system toward a new optimum operating point on a different branch may be the optimum course of action. For example, the management strategy for Lake Erie's fisheries introduces entirely new species and inter-species relationships and thus establishes a new ecosystem.

Fourth, an environmental change has implications for the ability of an ecosystem to respond to future environmental changes, even though the current change may have no immediate impact on the ecosystem. An example is the Fenitrothion spray of forests to control spruce budworm. This has no immediate impact, but interferes with the ability of the forest to regenerate itself in the face of other environmental changes. Similarly, forest fire suppression now appears to interfere with the ability of the forest to cope with fires at later times. The response of an ecosystem to environmental change, therefore, is a function of both the immediate environmental change and changes the ecosystem has been subjected to in the past. Historical environmental change can have both positive and negative implications for the ability of the system to cope with current changes.

3 Different Ecological Theories and Their State Variables

Having presented a framework for discussing ecosystem response to environmental change, we now turn our attention to different approaches to understanding ecosystems and the state variables, measures, and indicators that these approaches suggest. This discussion will give the reader a flavour of the variety of approaches being considered by ecologists as they struggle with understanding ecosystem development and response to environmental change. The discussion will be illustrative rather than comprehensive.

3.1 Inventories

The first step in understanding any ecological situation is to take inventory of the situation. This very time-consuming and expensive work is standard practice in environmental management and assessment. There are many different views about how to carry out such an inventory. (See Slocombe 1990 for a detailed discussion.) Two examples of the type of data collected are shown in Table 2. Such baseline data is crucial and usually requires much field work to acquire. Recent advances in Geographical Information Systems (GIS) allow such data to be tied to spatial locations. In essence, an inventory identifies how much of what is out there and where it is.

The reader should realize that for many ecological situations, this step uses most of the analyst's available resources. It is sobering to realize that for many ecological systems our knowledge is so limited that just an inventory often dramatically changes our understanding of

Table 2: Two Examples of Data Collected as Part of an Inventory
(after Slocombe and Woodley 1990)

Resource Description and Analysis Canadian Parks Service

CLIMATE: temperature, precipitation, wind, anomalies, topological effects.

WATER: drainage, hazards, quality, flow, timing, ice.

BEDROCK: geology, petrology, lithology, presence of fossil-bearing strata.

SURFICIAL MATERIALS: landforms, origins, types and descriptions of materials.

SOILS: soil orders and major soil characteristics.

FLORA: description of terrestrial and aquatic communities, distribution.

FAUNA: occurrence and distribution of terrestrial and aquatic species by habitat.

HIGH VISIBILITY AND SENSITIVE RESOURCES: number, distribution, status.

ARCHAEOLOGY-HISTORY: location, description and documentation of major events.

PAST AND PRESENT LAND USE: history of ownership, settlement, exploitation and transportation.

ECOLOGICAL HISTORY AND PROCESSES: past and future forces in evolution of park resources.

ABC Resource Survey Methodology

LEVEL 1 - RAW DATA

ABIOTIC

STRUCTURAL: landforms, slope, relief, drainage, soils.
FUNCTIONAL: geomorphic processes, deposition, avalanches, thermokarst action.

BIOTIC

STRUCTURAL: biotic land units, composition, coverage age, fire history.
FUNCTIONAL: areas of high plant-productivity, critical feeding, migration, winter habitats for wildlife.

CULTURAL

STRUCTURAL: past and present human land use.
FUNCTIONAL: corridors and activity nodes, spatial and temporal patterns of land use and cultural processes.

LEVEL 2 - INTERPRETIVE INDICES

ABIOTIC: uniqueness or representativeness, ecological importance, geomorphic hazard potential, terrain sensitivity.

BIOTIC: Faunal and community diversity, community uniqueness, faunal habitat dependence, vegetation recoverability, fire susceptibility.

CULTURAL: historical uniqueness or representativeness, archaeological importance, aesthetic or symbolic value, land use interactions, environmental impact.

the system. Most environmental management agencies focus on maintaining their inventories. Such information is purely descriptive and does not help us evaluate integrity or sustainability, except when rare species, resources, or ecosystems that need protection are identified, or changes in the inventory indicate significant changes in the system. In the latter case, the information is usually too late to prevent environmental degradation.

3.2 Population Studies

The next level of sophistication, after an inventory, is to study the populations of species in a system (see Pielou 1974). Such a study involves the collection and estimation of data and modelling. The issue is to predict the impact of environmental change on the number of individuals in a species. Environmental impact is judged in terms of changes in species numbers and the stability of the populations. The underlying assumption of this approach is that an ecosystem is made up of populations, and by understanding population dynamics one can understand ecosystems. This method is by far the most popular and most-funded approach for studying ecosystems. It is important when considering the effect of environmental change on individual species, but is not adequate for evaluating impacts at the ecosystem level.

One of the most widely used indicators in ecology is species diversity (see Pielou 1975; Grassle and others 1979). Measures of species diversity reflect the number of species in an ecosystem as well as the relative evenness of the population distributions. Preserving species diversity to conserve the unique characteristics of each species is an important principle for sustainable development. However in the early 1960s it was suggested that species diversity and ecological stability were intimately linked. Since that time, preservation of species diversity to ensure ecosystem stability has become the focus of much environmental management and many environmentalists. A thorough review of the diversity-stability hypothesis was undertaken by Goodman (1975). He showed conclusively that "clearly, the belief that more diverse communities are more stable is without support." Many good empirical studies that negate this hypothesis have also been published. However, the reader should be aware that species diversity is still the most popular indicator of ecosystem well-being, even though there is no scientific basis for representing it as such. Nevertheless, preserving species diversity is vital for ethical and aesthetic reasons, and to conserve the option values of unique biological compounds and structures.

Conservation biology is based largely on population studies. It is mission-oriented, the mission being the preservation of species from extinction. Two issues are of central importance: minimum viable population size and minimum ecosystem size. Minimum viable population size refers to the smallest number of individuals of a species required for the species to successfully continue reproducing. In general, the minimum viable population size is several thousand individuals. Minimum ecosystem size is the smallest area an ecosystem can occupy while still maintaining itself in its original state. Both are important numbers for the design of nature reserves to preserve species. These numbers represent estimates of critical thresholds in a catastrophe sense. Stay on one side of them and your system will continue to function. Cross them and catastrophic system collapse will occur. Thus, these numbers are important for assessing the sustainability of development from an ecological perspective.

3.3 Ecosystem Studies

Population studies represent a reductionist approach to understanding ecology, an approach based on the premise that the whole is the sum of the parts. However, we can be certain that complex systems exhibit emergent properties: 'The whole is greater than the sum of the parts'. The ecosystem approach competes with the population approach to ecology. Hence, fundamental to the ecosystem approach is the notion of an ecosystem, first put forward in detail by Odum (1953). The idea is that communities of species interact together to form a system, which must be studied as a whole; for example, a forest, a lake, or grasslands.

The ecosystem concept is much debated (see McIntosh 1985 or di Castri 1987 for illuminating discussions), and the reader should appreciate the profound implications of this controversy. Many do not consider the study of ecosystems as systems as a valid scientific activity. Most granting agencies, such as the Natural Sciences and Engineering Research Council, have not funded 'ecosystem studies'. Most scientific journals will not consider ecosystem studies for publication. Thus most ecosystem work is not formally funded. Much of the interesting work is found in books, management journals and the grey literature. What follows is a discussion of different approaches to studying ecosystems as systems.

3.4 Stability Theory

The concept of ecological stability has been the subject of much discussion[7]. The discussion was a first attempt to construct a framework similar to the one in Section 2. However, the notion of stability encompasses only the issue of whether an ecosystem can maintain its current optimum operating point (i.e., climax community) and does not explore whether the ecosystem will move to a new optimum operating point (i.e., a new climax). The problem with the notion of stability is that it is an equilibrium approach to system dynamics. It assumes that there is one equilibrium state that the system will attempt to maintain. This assumption is simply not valid for complex systems. The concept of ecological stability, however, was the first attempt to develop a notion of ecosystem integrity, and hence, measures of sustainability. It is still used extensively as the basis for environmental management, often hand in hand with preserving species diversity.

3.5 Stress-Response and Health

The idea underlying this approach is to parallel human health care (Rapport, Regier, and Hutchinson 1985; Barrett and Rosenberg 1981; Rapport 1989; Schaeffer and others 1988). Are ecosystems healthy or are they under stress? The key is to identify a set of diagnostic measures, such as blood pressure, heart rate, and temperature in humans. Once this is accomplished, then diagnostic techniques could be developed for evaluating ecosystem integrity and the continued sustainability of human interaction with the ecosystem. Table 3 shows various proposed sets of diagnostic indicators and how they would change with stress.

This approach is a very important step in developing our ability to evaluate ecological sustainability. It represents the current state of the art. It is important that the reader be aware of the limitations of this approach as the limitations will give the reader a strong sense of what the science of ecology, at its current stage of development, can tell us (see Schaeffer and others 1988, for a comprehensive analysis). The first problem is that there is no consensus on what constitutes stress or health. The first five chapters of Barrett and Rosenberg (1981) are devoted to discussions of how to define these terms. One person's stress is another person's response. Similarly, as Table 3 shows, we are not sure what the appropriate indicators of stress and health are. This uncertainty arises partly because of a lack of baseline data for ecosystems. We do not have the data sets to tell us what indicators for an ecosystem could be taken as analogous to normal human blood pressure or temperature. However, there are individuals who have sufficient understanding from their own site-specific observations to allow them to successfully identify indicators and values that represent health for specific local ecosystems.

Finally, there is a danger in pushing the analogy with human stress and health too far. Individual humans, except in unusual circumstances, have a very limited range of optimum operating points.[8] We do not normally exhibit bifurcations and flips in our operating point[9]. There is a danger that the stress-response and health analogy will focus our attention on maintaining a single optimum operating point for ecosystems. Thus, we might not be prepared for the kind of behaviour that involves changes in optimum operating points, whether it be a simple shift or a more dramatic bifurcation or flip (as discussed in Section 2). In addition, this analogy could lead to our unnecessarily, and perhaps inappropriately, constraining an ecosystem to a particular optimum operating point in the name of preserving its integrity. Such action could substantially interfere with the normal operation of an ecosystem (fire in a forest, for example) and degrade the ecosystem's potential for future development.

3.6 Flow Analysis

When E. Odum developed the concept of an ecosystem, it was in the context of the mass and energy flowing through the system. H.T. Odum developed this idea further and introduced the familiar energy flow diagrams (see Figure 5). Hannon, and later Patton and Ulanowicz, applied the mathematical framework of Leontief's and Augustinovics' economic input-output analysis to such flow networks. A set of measures were

[7] Appendix B is a short overview of these discussions and a good example of the pre-Newtonian state of ecology.

[8] Human blood pressure, heart rate, and temperature have narrowly defined ranges considered to be consistent with health.

[9] When we are ill, our temperature increases. This is a shift in optimum operating point. When children enter cold water, it triggers the diving reflex, which allows them to survive extended periods under water. This is a catastrophic change in optimum operating point. Similarly, if our food intake is substantially decreased for more than three weeks, our metabolism flips into a more efficient mode, thus making dieting harder.

Table 3: Comparison of Various Indicators of Stress in Ecosystems (after Slocombe and Woodley 1990)

Odum 1985

Energetics

1. Community respiration increases.
2. Production:respiration ratio becomes unbalanced.
3. Production:biomass ratio and respiration:biomass ratio increase.
4. Use of imported energy increases.
5. Unused primary production increases.

Nutrient Cycling

6. Increases nutrient turnover.
7. Horizontal transport increases, vertical cycling decreases.
8. Nutrient loss increases.

Community Structure

9. r-strategists increase.
10. Size of organism decreases.
11. Lifespan decreases.
12. Food chains shorten.
13 Species diversity decreases.

General System Level Trends

14. Ecosystems become more open.
15 Autogenic successional trends reverse.
16. Efficiency of resource use decreases.
17. Parasitism, etc., increases; mutualism, etc., decreases.
18. Functional properties are more robust than structural properties.

Rapport et al. 1985

1. Changes in nutrient cycling.
2. Changes in primary productivity.
3. Changes in species diversity.
4. System retrogression.
5. Changes in size distribution of species.
6. Increase in disease incidence.
7. Changes in the amplitude of fluctuations in component populations.

Schaeffer et al. 1988

1. Falling numbers of native species.
2. Overall regressive succession.
3. Changing standing crop biomass.
4. Changing relative amounts of energy flow to grazing and decomposer food chains.
5. Changes in mineral micronutrient stocks.
6. Changes in both the mechanisms of and capacity for damping undesirable oscillations.

Schlinder 1987
(for aquatic systems)

1. Phytoplankton production, standing crop, nutrient cycling and decomposition are not affected by pollutant stress.
2. Relative species abundance changes – opportunists favoured.
3. Taxonomic changes more important than diversity indices.
4. Species with short life cycles and poor dispersal are reduced. Greatest changes if there is no redundancy in the food web.
5. Life tables of sensitive species and keystone predators are best indicators.
6. Early stages of stress will not be detected by short-term toxicity studies or measures of ecosystem level processes.

Scheehan 1984
(for polluted ecosystems)

Individuals and Populations

1. Direct mortality or chronic accumulation causing death.
2. Sublethal impairment of behavior, physiology or function.
3. Altered growth and reproduction.

Ecosystem Structure and Dynamics

4. Reduction in population size and species extinction.
5. Loss of species with unique functions.
6. Decrease in species richness.
7. Changes in community composition and dominance patterns (indicator species).
8. Decreased species diversity.
9. Differences in taxonomy between healthy and polluted systems (similarity indices).
10. Changes in the ecosystem spatial structure.
11. Ecosystem properties of resilience, inertia, stability, constancy, elasticity, amplitude, hysteresis, and persistence.
12. Reversal of some aspects of succession with stress.

Ecosystem Functional Changes

13. Reduced decomposition, nutrient conservation, and primary productivity.
14. Increased energetic costs.
15. Alteration of cycles of essential nutrients, food web and functional regulation of ecosystem processes.

developed that characterize the energy-mass flow network in ecosystems (see Kay and others 1989; Ulanowicz 1986, for details). These measures describe important functional attributes of ecosystems, such as the amount of recycling being carried out, the use of nutrients and energy by the system, and the amount of throughput. As well, structural attributes such as the degree of interconnection between the components, a component's effective trophic level, and the number of trophic levels in the ecosystem, can be calculated. These have been successfully applied to a variety of ecosystems in the context of evaluating their development and health (Wulff and others 1990). An example of such an analysis is presented in Section 4.

Probably the ultimate empirical example of this approach is the Hubbard Brook Forested Ecosystem study conducted by Bormann and Likens (1979). In this study, data concerning all the inputs and outputs of four forested ecosystems (in various stages of development) were collected weekly over a 20-year period. The resulting picture of a forest is that of a giant water pump and filtration system. The bulk of the energy captured by forests goes to pumping water and recycling materials, not to growth. These results differ so radically from our ideas about forests that this study has not met with widespread acceptance.

The ability to further develop flow-analysis techniques is seriously limited by the data required. Adequate data sets for the evaluation of the various measures proposed do not exist. Furthermore, analysis of an ecosystem using this approach requires considerably more data collection than, for example, a population study. Although the analytical tools exist, considerable attention to the collection of appropriate data will be required before flow analysis will be viable as a way of evaluating ecosystem integrity.

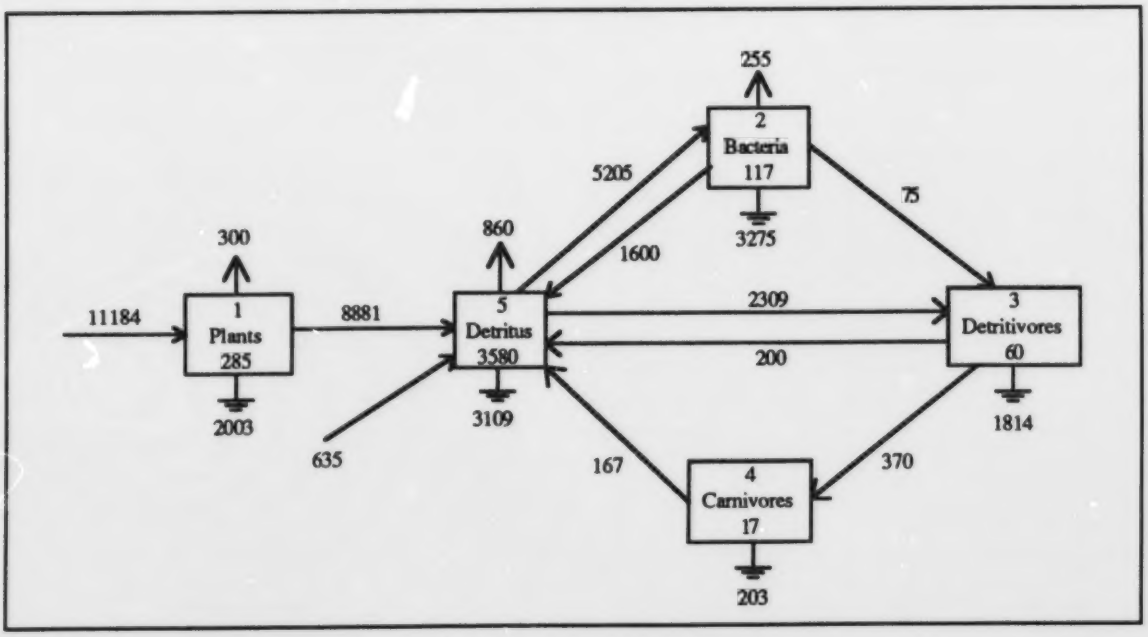


Figure 5: A simple weighted flow network representing the five components of the Cone Spring ecosystem. Flows of energy (kcal/m^2/yr) are indicated by the values that appear on the arcs, while standing stocks (kcal/m^2) are indicated inside the boxes that represent the nodes. There are two inputs that originate outside the system. The arcs that terminate outside the system represent exports of energy in a form that can be used by other systems. The special ground symbols (projecting from the bottom of each box) represent energy that is dissipated through respiration. This energy is lost by the system and is unusable by any system at the same scale. (after Kay and others 1989)

3.7 Summary

The flow-analysis approach to describing ecosystems is the antithesis of population studies. It deals with the ecosystem as a whole entity, a mass-and-energy flow system. It is consistent with non-equilibrium, self-organization analysis of other types of complex systems. Although it is clear to this author that ultimately our understanding of ecosystem development and integrity will follow from flow-analysis techniques, this does not diminish the importance of population studies. If we are to have sustainable development on this planet, understanding the ecology of individual species and maintaining their health are as important as maintaining the integrity of entire ecosystems.

Ecology is still a young science, struggling with basic concepts and questioning what to measure. We are only beginning to understand how ecosystems develop, and hence, what constitutes ecosystem integrity. Although the last decade has seen some encouraging advances in both theoretical and experimental ecology, it is too early to identify indicators of sustainability that would apply to all ecological situations.

4 Some Examples of Indicators and Their Use

Having left the reader with a sense that ecological indicators for sustainability may be a while in coming, this section presents some specific examples of indicators and their use in evaluating integrity and sustainability. The reader is cautioned, however, that the measures are specific to a geographical location or require data sets not generally available.

4.1 Some Sets of Indicators

One set of indicators arose from an Ecosystem Criteria Workshop whose task was "to identify meaningful ecosystem properties which can serve as operational guides for management of Green Bay or for Great Lakes ecosystems in general" (Harris and others, 1987). Table 4 shows the 18 indicators the workshop chose from a proposed set of 40. As can be seen from the tables, the chosen indicators look at mass-and-energy flow issues, population issues, and human aesthetic and use issues. The reader can probably appreciate that just collecting the baseline data necessary to use these measures is a formidable task.

A second example is the Indicator of Biological Integrity (IBI) measure proposed by Karr and others (1986). Tables 5 and 6 describe this indicator. It is primarily a population-based indicator. It is meant to assess fish communities in streams in the midwestern United States. Steedman (1988) has modified this measure for streams in southern Ontario. Although this measure has been used quite successfully to evaluate integrity, it is limited geographically and to fish.

Schaeffer and others (1988) identified 11 criteria for measures of ecosystem health – criteria that apply equally to ecological indicators of sustainability. Some of these criteria relate to the mathematical well-behavedness of the measures, something often neglected in the biological sciences. As well, the measures should:

a) not depend on the presence, absence, or condition of a single species;

b) not depend on a census or inventory of a large number of species;

c) reflect our knowledge of normal succession;

d) not be a single number; and

e) be related and hierarchically appropriate for use in ecosystems.

The reader would be well advised to keep these criteria in mind when evaluating ecological indicators of sustainability.

4.2 Comparative Ecosystem Study

Wulff and Ulanowicz (1989) undertook a comparison of the Baltic Sea and the Chesapeake Bay using flow-analysis techniques. Figure 6 shows the carbon flow diagrams for the two ecosystems. The Baltic has a lower species diversity, its primary production is one-third that of Chesapeake, its total system throughput is 20% of Chesapeake's, and it is 33 times bigger than Chesapeake Bay. These traditional measures indicate that of the two ecosystems, the Baltic is in poorer shape than the Chesapeake. However,Wulff and Ulanowicz found ". . . that the Baltic was trophically more efficient and possesses a more highly structured array of recycling loops than does the Chesapeake. . . these characteristics suggest that the Chesapeake is subject to a more intense set of stresses than the Baltic."

In particular, Wulff and Ulanowicz found that, overall, the effective trophic level of each species was higher in the Baltic (see Figure 7). Species tend to be as highly trophically situated as possible, as this will result in the highest energy density in the food they catch. Lower effective trophic levels are, therefore, associated with stress. Overall, the Baltic has more trophic levels, 7 versus 6, than the Chesapeake (see Figure 9). These results indicate that the Chesapeake is more stressed than the Baltic.

The trophic efficiencies in the Baltic are much higher. More of the captured energy is passed up the food web. Also the ratio of energy input to the grazing chain from detritus, relative to input from herbivory is 4:1 for the Chesapeake, versus 1.5:1 in the Baltic. Finally, the number of cycles is higher in the Baltic (20 versus 14), and the length of the cycles is longer in the Baltic (see Figure 8). Cycles are food-web pathways that recycle food back to their starting point to be reused again. More and longer cycles allow for more complete use of food. These last two observations indicate more complete usage of the available resources in the Baltic. Overall, this set of observations indicates that the Baltic is making better use of its captured resources. Its food web is more organized than the Chesapeake's.

4.3 Evaluating an Ecosystem Affected by Humans

This is an example of evaluating the impact of human development on an ecosystem.[10] The ecosystem in question is a marsh gut adjacent to Crystal River, Florida.[11] The stressed ecosystem is exposed to hot water effluent from a nuclear power station. The stress is a water-temperature increase of approximately 6°C. The control ecosystem is not exposed to the effluent, but is otherwise exposed to the same environmental conditions.

Table 7 summarizes a set of ecosystem indicators. The input-output measures indicate various aspects of the flows through the ecosystem. In absolute terms, all the flows have dropped in the stressed ecosystem. Overall, the drop in flows is about 20%; in particular, the imported flows (that is, the resources available for consumption) drop by 18% and the total system throughput (TST – the total flow activity in the system) drops by 21%. The biomass drops by about 35%. The implication of these numbers is that the stress has resulted in the ecosystem shrinking in size, both in terms of biomass and its consumption of resources.

If the flows are scaled by the import to the ecosystem from the outside, the resulting numbers indicate how well the ecosystem is making use of the resources it does capture. The percentage changes in these scaled flow rates reveals that, in total, the stressed ecosystem is exporting material more quickly than the control ecosystem. Looking at the living side, the big changes are an increase in respiration and a decrease in flow to the detritus. There is a small decrease in exports. These changes indicate that the consumed resources are being more effectively used by the living components. It also indicates the species are stressed.[12] Looking at the detritus, a different picture emerges. There is a slight increase in production, a small decrease in respiration, and a very large increase in exports. This analysis of the scaled flow rates indicates that the living components are somewhat stressed but, more important, there is a big breakdown in the ability of the ecosystem to recycle material through the detritus, and thus retain its resources.

Examining the food web data further confirms this. The number of cycles in the stressed ecosystem is 51% of the number in the control. Furthermore, these cycles are shorter in overall length (see Figure 10). Looking at the effective grazing chain (Figure 11), the number of trophic levels has not changed, but the trophic efficiencies of compartments III and IV have changed dramatically, as has the flow to the top trophic levels. These are all indicators of a stressed ecosystem.

The impact of the effluent from the power station has been to decrease the size of the ecosystem and its consumption of resources, while affecting its ability to retain the resources it has captured. In short, the impacted ecosystem is smaller, has lower trophic levels, recycles less, and leaks. All of these are signs of disorganization and a step backward in development.

[10] This evaluation is at the ecosystem level only. To completely evaluate the impact, it is necessary to examine species level impacts (the next level down) and the broader regional impacts (the next level up).

[11] The data was collected by Homer and Kemp but never published.

[12] A stressed individual will increase respiration and decrease excrement.

Table 4: Ecosystem Indicators for Green Bay (after Harris and others 1987)

1. Input-output property of system
   - Nutrients
   - Particulate materials
   - Energy
2. Residence time
   - Water (circulation and flushing rates)
   - Toxics
   - Particulates
   - Nutrients (nutrient cycling)
3. Sedimentation rates
   - Organic carbon flux
   - Toxics flux
   - Hypolimnetic O_2 levels
4. Primary production
   - Macrophytes versus phytoplankton
   - Algal biomass and food quality (for zooplankton grazing)
5. Secondary production and biomass
   - Detrital food chain
   - Zooplankton food quality (for planktivore grazing)
   - Fish mortality and natality
   - Reproduction and growth rate of preferred fish species
6. Trophic structure
   - Length of food chain
   - Limitation of energy flow between one level and another (depends on quality of food at each level)
   - Species size distribution and species diversity (affects quality of food available to next trophic level)
   - P/B level can be specific to trophic levels and turnover rates
7. Ratio of pelagic versus benthic biomass
8. Carbon transfer efficiency
   - Applies at each level of food chain (however at differing levels of importance)
   - Especially important in detrital food chain
9. Connectivity
   - Within and between trophic level: subset of trophic structure (above)
   - Affects stability of system
10. Predatory prey species abundance
    - Applies at each trophic level
11. Turn-over rate of slowest variable in system
    - Keystone species are at top of food chain and have longest life cycle
    - Physicochemical reactions (for example, remineralization of sediment nutrients) are important
12. Conditions of top piscivore
    - Top predator growth rate, productivity
    - Body burden of toxic chemicals
13. Key indicator species
    - Representative of most selective species at various trophic levels
    - Integrators of environmental quality
14. Toxic effects
    - Inhibition on productivity (primary and secondary)
    - Organismal physiology
    - Bioaccumulation factors
    - Fish body burdens
15. Variability of system parameters over time
    - Flux in physical and chemical conditions
    - Community functions
    - Species functions
    - Pollution levels
16. Uses and perception of uses
    - Aesthetics
    - Human values (economic basis)
17. Length of time for a system to return to functional (healthy state)
18. Ratio of fossil to solar fuel in system (needed for estimate of true or unsubsidized net productivity)

Table 5: Components of the Indicator of Biological Integrity (after Karr and others 1986)

Category	Metric
Species richness and composition	1. Total number of fish species. 2. Number and identity of darter species. 3. Number and identity of sunfish species. 4. Number and identity of sucker species. 5. Number and identity of intolerant species. 6. Proportion of individuals as green sunfish. 7. Proportion of individuals as omnivores. 8. Proportion of individuals as insectivorous cyprinids. 9. Proportion of individuals as piscivores (top carnivores).
Fish abundance and condition	10. Number of individuals in sample. 11. Proportion of individuals as hybrids. 12. Proportion of individuals with disease, tumors, fin damage, and skeletal anomalies.

Table 6: Total IBI Scores, Integrity Classes, and Attributes of the Classes (after Karr and others 1986)

Total IBI score (sum of the 12 metric ratings)	Integrity Class	Attributes
58-60	Excellent	Comparable to the best situations without human disturbance; all regionally expected species for the habitat and stream size, including the most intolerant forms, are present with a full array of age (size) classes; balanced tropic structure.
48-52	Good	Species richness somewhat below expectation, especially because of the loss of the most intolerant forms; some species are present with less than optimal abundances or size distributions; trophic structure shows some signs of stress.
40-44	Fair	Signs of additional deterioration include loss of intolerant forms, fewer species, highly skewed trophic structure (e.g., increasing frequency of omnivores and green sunfish or other tolerant species); older age classes of top predators may be rare.
28-34	Poor	Dominated by omnivores, tolerant forms, and habitat generalists; few top carnivores; growth rates and condition factors commonly depressed; hybrids and diseased fish often present.
12-22	Very poor	Few fish present, mostly introduced or tolerant forms; hybrids common; disease, parasites, fin damage, and other anomalies regular.
	No fish	Repeated sampling finds no fish.

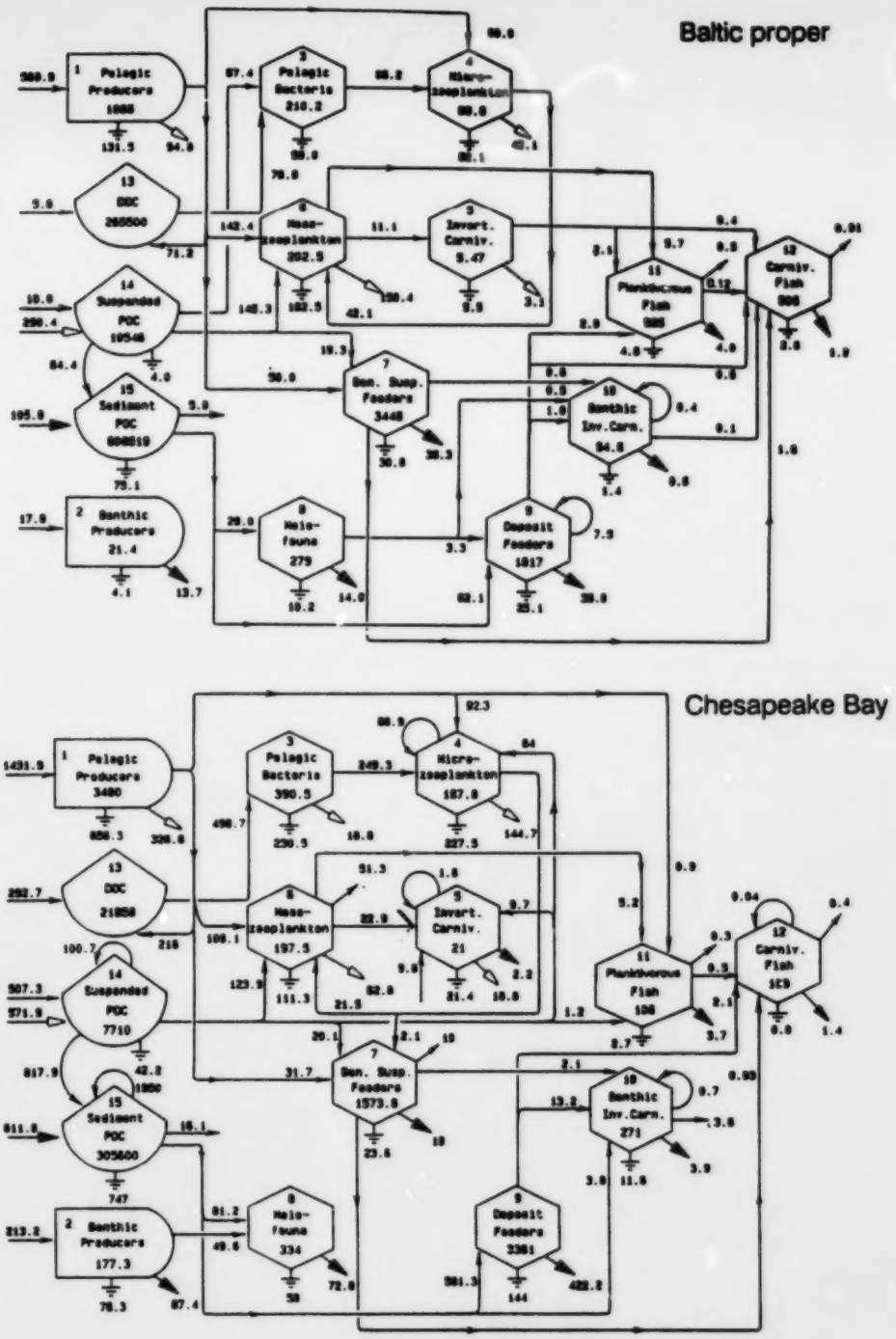


Figure 6: Carbon flow diagrams for the Baltic Sea and Chesapeake Bay (mg m^{-2} day^{-1}, after Wulff and Ulanowicz 1989)

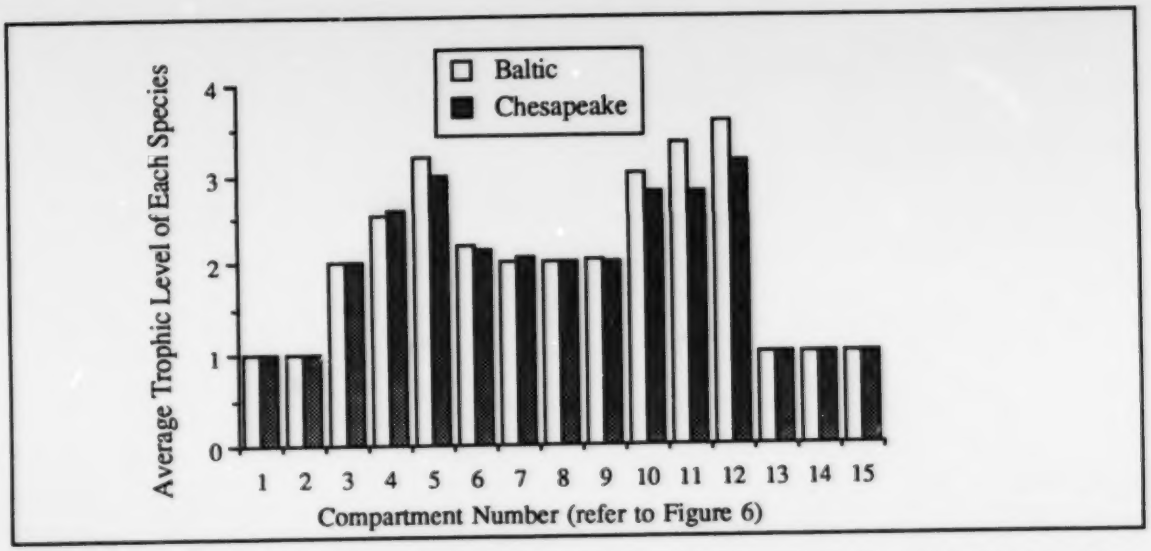


Figure 7: Effective Trophic Level of Each Species in the Baltic Sea and Chesapeake Bay

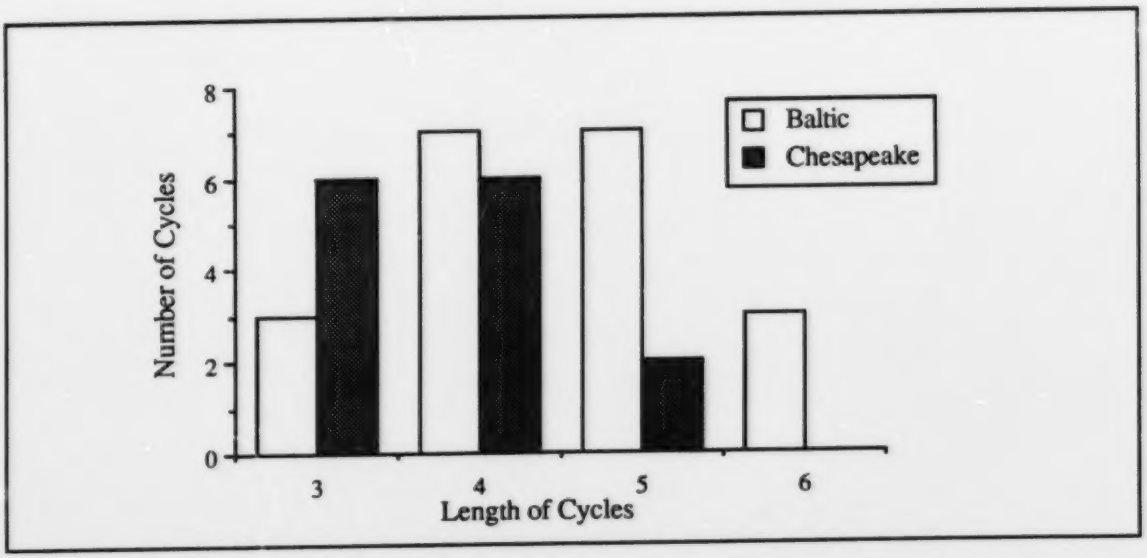


Figure 8: Number of Cycles of the Given Cycle Length for the Baltic Sea and Chesapeake Bay

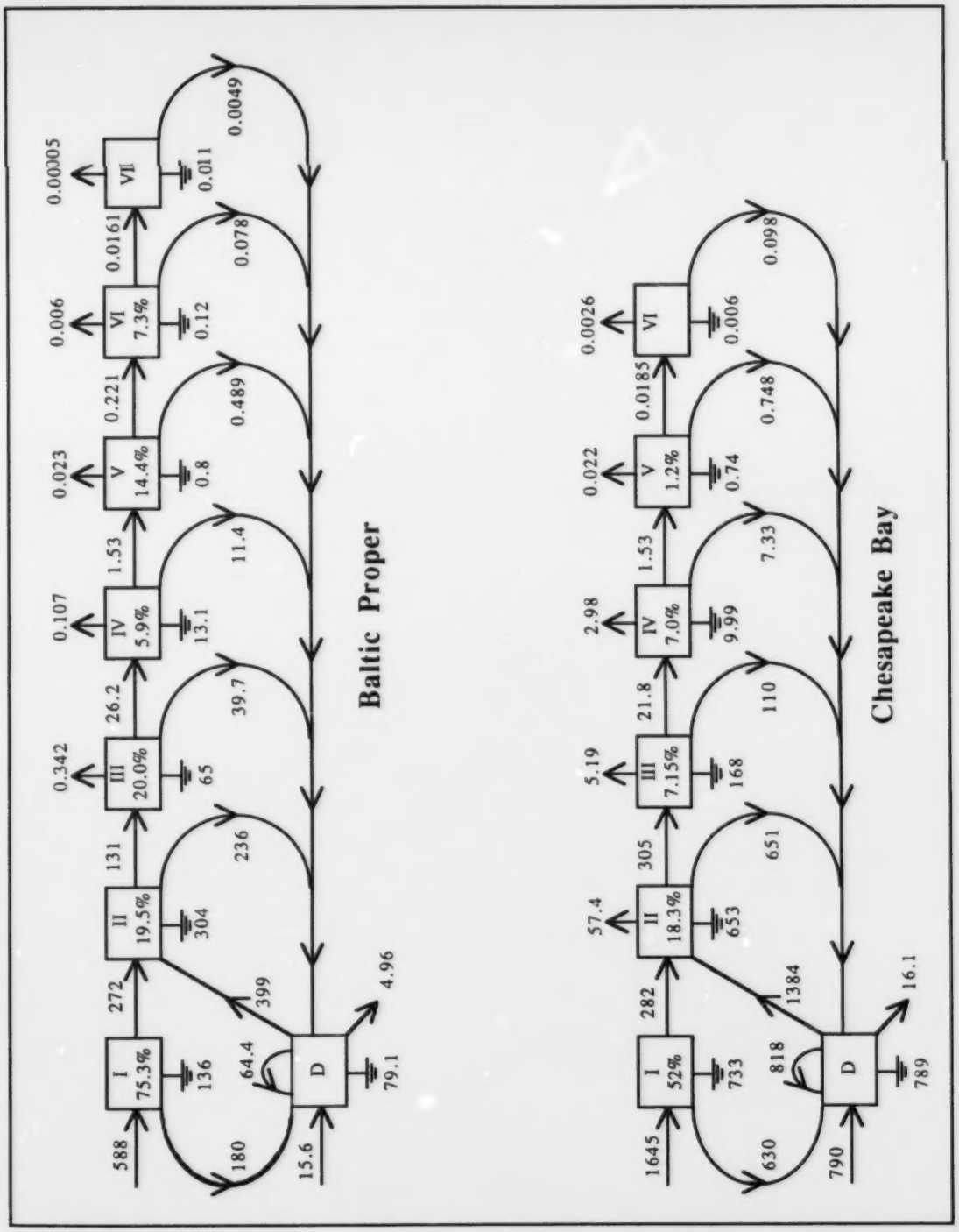


Figure 9: **The effective grazing chain and detrital recycling for the Baltic Sea and Chesapeake Bay.** The percentages in boxes are the overall trophic efficiencies. (after Wulff and Ulanowicz)

Table 7: Ecosystem Indicators for the Crystal River Marsh Gut Ecosystems

Crystal River	mg/sq m/day				Scaled by Import		
	Control	Stressed	Δ%		Control	Stressed	Δ%
Biomass	1,157,136	755,213	34.73%		157.50	125.49	20.32%
Total I/O							
Imports	7,347	6,018	18.09%		1	1	0.00%
TST*	22,768	18,055	20.70%		3.10	3.00	3.19%
Production	3,292	2573.9	21.82%		0.45	0.43	4.56%
Exports	952	872	8.37%		0.13	0.14	-11.86%
Respiration	6,400	5,148	19.57%		0.87	0.86	1.80%
Living I/O							
Production	400	326	18.39%		0.05	0.05	0.37%
Exports	316	253	19.93%		0.04	0.04	2.24%
Respiration	3,566	3,078	13.69%		0.49	0.51	-5.37%
Detritus	5,726	4,315	24.64%		0.78	0.72	7.99%
Detritus I/O					**Scaled by Detritus Input**		
Input from Living	5,726	4,315	24.64%		1.00	1.00	0.00%
Production	2,893	2,248	22.29%		0.51	0.52	-3.11%
Exports	636	619	2.64%		0.11	0.14	-29.19%
Respiration	2,834	2,070	26.96%		0.49	0.48	3.09%
Food Web							
Cycles	142	69	51.41%				
Trophic Levels	5	5	0.00%				

* Total System Throughput is a measure of the total flow activity in the ecosystem.

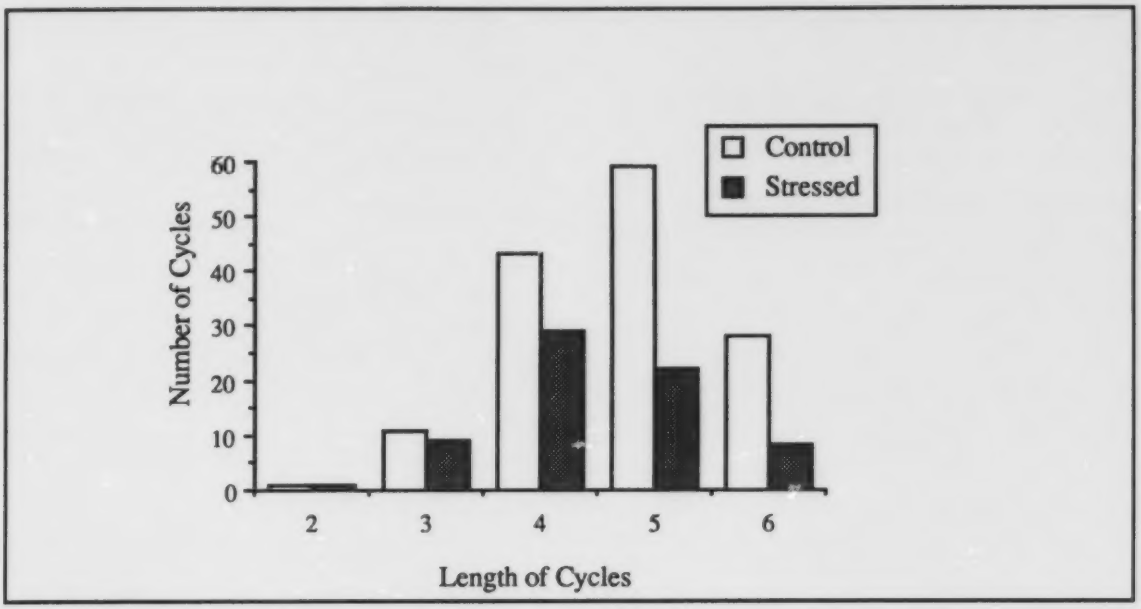


Figure 10: The Number of Cycles of Each Cycle Length for the Crystal River Ecosystems

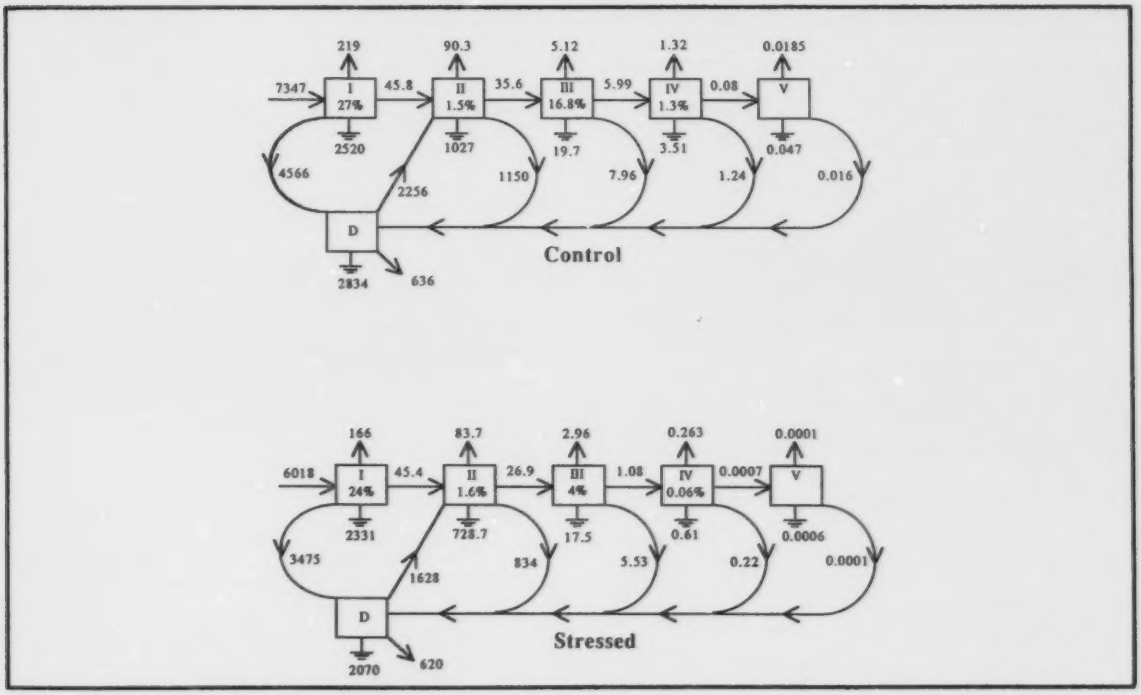


Figure 11: The effective grazing chain and detrital recycling for marsh guts near Crystal River, Florida. The percentages in boxes are the overall trophic efficiencies.

5 Concluding Remarks

Ecology is a young discipline, particularly in the study of ecosystems. As such, it is premature to expect to be able to identify a universal set of indicators of integrity, and hence, to have a formal quantitative methodology for evaluating ecological sustainability in all situations. The analyst, therefore, must tailor measures and methodology for the particular ecological sustainability issue being examined. In all cases, only qualitative assessments are possible. In all cases prudence should be the guiding principle. To err on the side of caution is to avoid crossing a catastrophe threshold and permanently losing a species or ecosystem. In this vein, Holling's Adaptive Environmental Management (1978) is well worth consideration.

An important action that Canada could undertake would be the formal support of theoretical ecology at the ecosystem level and the collection of baseline data for ecosystems. Lack of baseline data severely constrains our ability to develop indicators of ecosystem integrity and health.

Without these or a good understanding of ecosystem dynamics, our ability to make sound decisions about what constitutes sustainable development from an ecological perspective will continue to be impaired.

I would add a final caveat. By their nature, dissipative structures exhibit surprising behaviour. This behaviour cannot be *a priori* predicted and may be catastrophic. No matter how much knowledge and understanding we have, we will always be subject to surprise when we observe ecosystems. Therefore, any human systems that are meant to deal with ecosystems (or any dissipative systems) must be adaptive in their response and able to cope with surprises.

Acknowledgements

The author wishes to thank Dr. Eric Schneider, Chief Scientist, National Ocean Service, (U.S.A.) and Professor Bob Ulanowicz of Chesapeake Biological Laboratory, University of Maryland for many, many discussions on this subject and for providing many data sets. Section 2 and the appendices are drawn from a paper originally presented to the IJC/GLFC workshop on ecosystem integrity, which is currently in press with Environmental Management. Reviewers of this paper, in particular Dr. T. Fontaine, Director, Water Quality Division, South Florida Water Management District, are greatly acknowledged. Two of my students, Mr. S. Woodley and Dr. S. Slocombe contributed substantially through unpublished literature reviews. Members and staff of CEAC, particularly Dr. S. Conover and Mr. J. Potvin, made many suggestions that substantially improved this paper.

References

Barrett, G.W., and R. Rosenberg. 1981. *Stress Effects in Natural Ecosystems*. J. Wiley & Sons.

Bormann, F., and G. Likens. 1979. *Pattern and Process in a Forested Ecosystem*. Springer-Verlag.

Cairns, J.P., and K. Dickson. 1977. Recovery of Streams from Spills of Hazardous Materials. In *Recovery and Restoration of Damaged Ecosystems*, by J.P. Cairns, K. Dickson, and E. Herricks. University Press of Virginia, 24-42.

DeAngelis, D.L., et al. 1989. Nutrient Dynamics and Food Web Stability. *Annual Review of Ecology and Systematics*. 20:71-95.

di Castri, F. 1987. The Evolution of Terrestrial Ecosystems. In *Ecological Assessment of Environmental Degradation, Pollution and Recovery*, by O. Ravera (ed.). Elsevier Science, 1-30.

Goodman, D. 1975. The Theory of Diversity-Stability Relationships in Ecology. *Q. Rev. Biol.* 50(3):237-366.

Granero-Porati, M.I., R. Kron-Morelli, and A. Porati. 1982. Random Ecological Systems with Structure: Stability-Complexity Relationship. *Bulletin of Mathematical Biology* 44(1):103-117.

Grassle, J.F., et al. (eds.) 1987. *Ecological Diversity in Theory and Practice*. International Co-Operative Publishing House. Statistical Ecology Series, 6:275.

Harris, H.J., et al. Coupling Ecosystem Science with Management: A Great Lakes Perspective from Green Bay, Lake Michigan, USA. *Env. Mgmt.* 11(5):619-625.

Harte, J., and D. Levy. 1975. On the Invulnerability of Ecosystems Disturbed by Man. In *Unifying Concepts in Ecology*, by W.H. van Dobben and R. Lowe-McConnell (eds.). The Hague: Dr. W. Junk B.V. Publishers.

Hill, A.R. 1975. Ecosystem Stability in Relation to Stresses Caused by Human Activities. *Canadian Geographer* 19:206-220.

Hirata, H., and T. Fukao. 1977. A Model of Mass and Energy Flow in Ecosystems. *Mathematical Biosciences* 33.

Holling, C.S. 1978. *Adaptive Environmental Assessment and Management*. J. Wiley & Sons.

Holling, C.S. 1973. Resilience and Stability of Ecological Systems. *Annual Review of Ecology and Systematics* 4.

Holling, C.S. 1986. The Resilience of Terrestrial Ecosystems: Local Surprise and Global Change. In *Sustainable Development in the Biosphere*, by W.M. Clark and R.E. Mann (eds.). Oxford University, 292-320.

Huseyin, K. 1977. The Multiple-Parameter Stability Theory and Its Relation to Catastrophe Theory. In *Problem Analysis in Science and Engineering*, by F.H. Branin and K. Huseyin (eds.). Academic Press, 229-255.

Huseyin, K., and V. Manadi. 1980. On the Instability of Multiple-Parameter Systems. In *Theoretical and Applied Mechanics*, by F.P.J. Rimorott and B. Tabarrok (eds.). Amsterdam: North-Holland, 281-294.

Jones, D.D. 1975. The Applications of Catastrophe Theory to Ecological Systems. In *New Directions in the Analysis of Ecological Systems*, by G.S. Innis (ed.). 5:133-148. (Simulation Council Proceedings Series.)

Karr, James R., et al. 1986. *Assessing Biological Integrity in Running Waters: A Method and Its Rationale*. Illionois Natural History Survey. Special Publication 5, 28.

Kay, James. 1984. *Self-Organization in Living Systems*. Ph.D. thesis, Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.

Kay, James J., Lee Graham, and Robert E. Ulanowicz. 1989. A Detailed Guide to Network Analysis. In *Network Analysis in Marine Ecosystems*, by F. Wulff, J.G. Field, and K.H. Mann (eds.). Springer-Verlag. 32:15-61. (Coastal and Estuarine Studies.)

Kay, J.J. 1989. A Thermodynamic Perspective of the Self-Organization of Living Systems. In *Proceedings of the 33rd Annual Meeting of the International Society for the System Sciences*, by P.W.J. Ledington (ed.). Edinburgh, 3:24-30.

Lewotin, R. 1969. The Meaning of Stability. In *Diversity and Stability in Ecological Systems*, by G.M. Woodwell and H.H. Smith (eds.). Brookhaven National Laboratories Symposium 22.

Ludwig, D., D.D. Jones, and C.S. Holling. 1978. Qualitative Analysis of Insect Outbreak Systems: The Spruce Budworm and the Forest. *Journal of Animal Ecology* 44:315-332.

Lugo, A.E., G. Cintron, and C. Goenaga. 1981. Mangrove Ecosystems under Stress. In *Stress Effects in Natural Ecosystems*, by G.W. Barrett and R. Rosenberg (eds.). J. Wiley & Sons, 129-153.

Margalef, R. 1975. Diversity, Stability, and Maturity in Natural Ecosystems. In *Unifying Concepts in Ecology*, by W.H. van Dobben and R. Lowe-McConnell (eds.). The Hague: Dr. W. Junk B. V. Publishers, 151-160.

May, R. 1974. General Introduction. In *Ecological Stability*, by M. Usher and M. Williamson (eds.). Chapman & Hall.

May, R. 1985. Thresholds and Break Points in Ecosystems with a Multiplicity of Stable Points. *Nature*. 269:471-477.

McIntosh, R. P. 1985. *The Background of Ecology: Concept and Theory*. Cambridge University Press.

Nelson-Smith, A. 1977. Recovery of Some British Rocky Seashores from Oil Spills and Cleanup Operations. In *Recovery and Restoration of Damaged Ecosystems*, by J.P. Cairns, K. Dickson, and E. Herricks. University Press of Virginia, 191-207.

Nicolis, G., and I. Prigogine. 1989. *Exploring Complexity*. W.H. Freeman.

Nicolis, G., and I. Prigogine. 1977. *Self-Organization in Non-Equilibrium Systems*. J. Wiley & Sons.

Odum, E.P. 1985. Trends Expected in Stressed Ecosystems. *BioScience* 35(7):419-422.

Odum, E.P. 1953. *Fundamentals of Ecology*. Saunders.

Odum, E.P. 1969. The Strategy of Ecosystem Development. *Science* 164:262-270.

Orians, G.H. 1975. Diversity, Stability, and Maturity in Natural Ecosystems. In *Unifying Concepts in Ecology*, by W.H. van Dobben and R. Lowe-McConnell (eds.). The Hague: Dr. W. Junk B.V. Publishers, 139-150.

Pielou, E.C. 1975. *Ecological Diversity*. J. Wiley & Sons.

Pielou, E.C. 1974. *Population and Community Ecology*. Gordon & Breach.

Preston, F. 1969. Diversity and Stability in the Biological World. In *Diversity and Stability in Ecological Systems*, by G.M. Woodwell and H.H. Smith (eds.). Brookhaven National Laboratories Symposium 22.

Prigogine, I., G. Nicolis, and A. Babloyantz. 1972. Thermodynamics of Evolution. *Physics Today* 23(11):23-28.

Prigogine, I., G. Nicolis, and A. Babloyantz. 1972. Thermodynamics of Evolution. *Physics Today* 23(12):38-44.

Rapport, D.J., H.A. Regier, and T.C. Hutchinson. 1985. Ecosystem Behaviour under Stress. *Am. Nat.* 125:617-640.

Rapport, D.J. 1989. What Constitutes Ecosystem Health. *Perspectives in Biology and Medicine* 33(1):120-132.

Robinson, J.V. and W.D. Valentine. 1979. The Concepts of Elasticity, Invulnerability and Invadability. *Journal of Theoretical Biology* 81:91-104.

Rutledge, R.W. 1974. Ecological Stability: A Systems Theory Viewpoint. *Electrical Engineering*. Oklahoma State University.

Rutledge, R.W., B.L. Basore, and R.J. Mulholland. 1976. Ecological Stability. *Journal of Theoretical Biology* 57:355-371.

Schaeffer, David, J., E. Herricks, and H. Kerster. 1988. Ecosystem Health: I. Measuring Ecosystem Health. *Env. Mgmt.* 12(4):445-455.

Schlinder, J.E. 1987. Detecting Ecosystem Responses to Anthropogenic Stress. *Can. J. Fish. Aquat. Sci.* 44:6-25.

Sharitz, R., and J.W. Gibbons. 1981. Effects of Thermal Effluents on a Lake: Enrichment and Stress. In *Stress Effects in Natural Ecosystems*, by G.W. Barrett and R. Rosenberg (eds.). J. Wiley & Sons, 243-259.

Sheenan, P.J. et al. (eds.).1984. *Effects of Pollution at the Ecosystem Level*. J. Wiley & Sons.

Shure, D.J., and E.J. Hunt. 1981. Ecological Response to Enrichment Perturbation in a Pine Forest. In *Stress Effects in Natural Ecosystems*, by G.W. Barrett and R. Rosenberg (eds.). J. Wiley & Sons, 103-114.

Slocombe, S., and S. Woodley. 1990. *Environmental Monitoring in National Parks: A Bibliographic Review*. Waterloo, Ontario: Heritage Resource Center, University of Waterloo.

Steedman, R. 1988. Modification and Assessment of an Index of Biotic Integrity to Quantify Stream Quality in Southern Ontario. *Can. J. Fish. Aquat. Sci.* 45(1):45-501.

Stokes, Pamela. 1984. Clearwater Lake: Study of an Acidified Lake Ecosystem. In *Effects of Pollutants at the Ecosystem Level*, by P.J. Sheehan et al. (eds.). SCOPE Series, 22:229-253. J. Wiley & Sons.

Thom, R. 1969. Topological Models in Biology. *Topology* 8.

Ulanowicz, R.E. 1979. Complexity, Stability, and Self-Organization in Natural Communities. *Oecoligia*. 43:295-298.

Ulanowicz, R.E. 1980. An Hypothesis on the Development of Natural Communities. *Journal of Theoretical Biology* 85:223-245.

Ulanowicz, R.E. 1986. *Growth and Development: Ecosystem Phenomenology*. Springer-Verlag.

Usher, M., and M. Williamson (eds.). 1974. *Ecological Stability*. Chapman & Hill.

van Voris, P., R.V. O'Neill, W.R. Emanuel, and H.H. Shugart. 1980. Functional Complexity and Ecosystem Stability. *Ecology* 61(6):1352-1360.

Walker, B.H., D. Ludwig, C.S. Holling, and R.M. Peterman. 1981. Stability of Semi-Arid Savanna Grazing Systems. *Journal of Ecology* 69:473-498.

Weinberger, P., R. Greenhalgh, and R.P. Moody. 1981. Fenitrothion as a Wide-Ranging Perturbation Factor in the Environment. In *Stress Effects in Natural Ecosystems*, by G.W. Barrett and R. Rosenberg (eds.). J. Wiley & Sons, 155-176.

Wu, L. 1974. On the Stability of Ecosystems. In *Ecosystem Analysis and Prediction*, by S.A. Levin (ed.). Society for Industrial and Applied Mathematics.

Wulff, F., J.G. Field, and K.H. Mann (eds.). 1989. *Network Analysis in Marine Ecosystems* 32. Springer-Verlag. (Coastal and Estuarine Studies.)

Wulff, F., and R.E. Ulanowicz. 1989. A Comparative Anatomy of the Baltic Sea and Chesapeake Bay Ecosystems. In *Network Analysis in Marine Ecosystems*, by F. Wulff, J.G. Field, and K.H. Mann (eds.). Coastal and Estuarine Studies Series. 32:232-258. Springer-Verlag.

Appendix A: Some Systems Notions

Throughout this paper some notions from systems theory are used. These are described in this appendix for those readers unfamiliar with them.

A **state variable** is a variable that describes some aspect of the system we are interested in. In population modelling, the number of individuals of a species would be the state variable. E.P. Odum (1969) identified a now famous set of ecosystem state variables and how they change with succession. Examples of these are photosynthesis and respiration rates, net productivity, total biomass, and species diversity. Ulanowicz (1986) has identified a set of variables that describe the state of an ecosystem's food web. Some of these are ascendency, number of cycles, cycling index, and effective trophic levels. Which state variables are looked at depends on the questions posed by the researcher.

A **state space** is a multidimensional space whose axes are the state variables. In a predator-prey system, the state variables would be the population numbers for each species and the two-dimensional space, with the number of predators on one axis and the number of prey on the other, would be the state space. There would be a curve (a **path** in state space) that describes the relationships between the predator and the prey. (Figures A-1a and A-1b are examples.)

For a given set of forces acting on a system, there will be at least one point in state space where the forces are balanced. This is known as the **equilibrium** point. For example, the equilibrium point for a population is the point where the mortality and birth rates balance. The issue of importance is the **stability** of the equilibrium point. That is, is the system able to stay in equilibrium? Consider a cone that has a very narrow blunt top. If it is placed upside down on its top, then a small disturbance will cause it to fall over. On the other hand, if it is placed with its

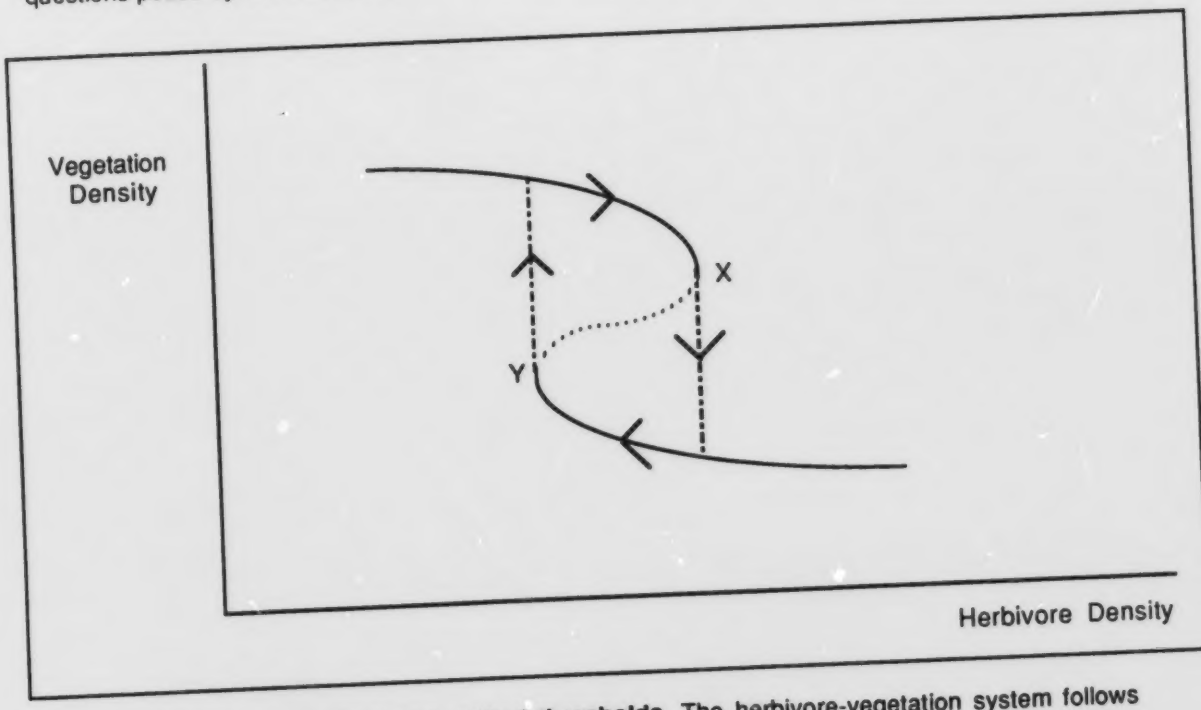


Figure A-1a: Catastrophes and critical thresholds. The herbivore-vegetation system follows the equilibrium path through its state space as indicated by the arrows. At point X the equilibrium path becomes unstable and the system drops from the upper solid curve to the lower solid curve. At point Y the system is again unstable and moves from the lower to the upper solid curve.

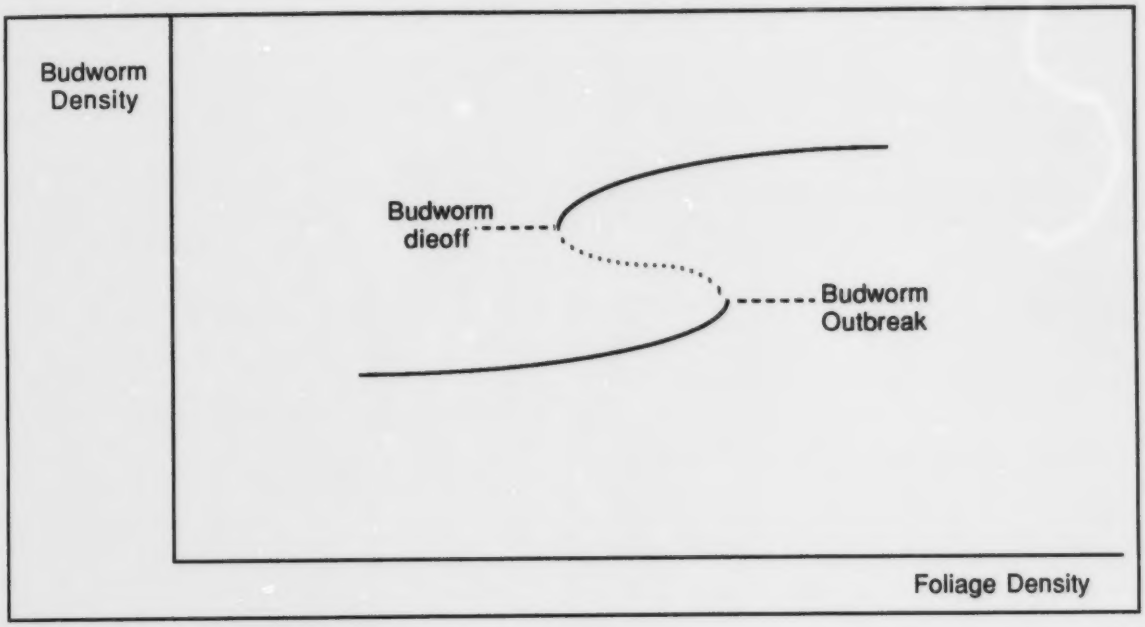


Figure A-1b: A Similar Situation for Spruce Budworm

top up and its broad base down, only a very large disturbance will cause the cone to topple over. In the former case, the equilibrium is said to be unstable; in the latter, stable. In a strict mathematical sense, an equilibrium point is stable if, after a disturbance, the state of the system returns to the equilibrium point.

Another possibility is that the state of the system does not return to the equilibrium point after a disturbance but oscillates about it with a maximum amplitude. Consider a perfect pendulum. The equilibrium point is at the bottom of the swing. The system oscillates about this point with a maximum amplitude after it has been disturbed. In the case of a real pendulum, it eventually comes to rest at the equilibrium point. Both the ideal and real pendulum are considered stable.

These two types of stability are mathematically defined by the **Lyapunov stability** criteria. (See Lewontin (1969) and Harte and Levy (1974) for a review of this theory as applied to ecology.) The key question is whether the state of the system will return to the equilibrium point or oscillate about it when the system is disturbed, or if the state of the system is permanently moved to another point in state space (the cone falls over). It was through a Lyapunov stability analysis of thermodynamic systems, with entropy production as the state variable, that Prigogine discovered the self-organizing phenomena for which he was awarded the Nobel prize.

The **optimum operating point** for an ecosystem is an equilibrium point in state space that represents a balance between the forces acting on the ecosystem. In the real world, the environment is not static. The forces acting on an ecosystem are constantly changing. Therefore, the equilibrium point is constantly changing. For the purpose of discussion in this paper, the optimum operating point has been treated as being stationary. In reality, it is constantly changing and would be more realistically represented by a distribution in time. This distribution would reflect the distribution of environmental parameters.

Notwithstanding this variability, it is possible that the ecosystem has cyclic stability much like a pendulum. Holling (1986) has shown this to be the case for some pest outbreaks in forested ecosystems that happen with a fixed frequency. The forested ecosystem swings between maximum foliage just before an outbreak and minimum foliage just before the outbreak ceases. Another example is ecosystems driven by phytoplankton blooms.

A very important systems notion is that of a catastrophe. Catastrophe theory was brought to prominence by Rene Thom (1969), and an analytical basis for it was discovered by Huseyin (1977, 1980). The importance of catastrophe theory is that it shows how systems can exhibit behaviour that is discontinuous and occurs without warning. Usually the phenomenon is very dramatic.

A simple example is shown in Figure A-1a. As the herbivore population increases, the vegetation decreases; more is eaten. Eventually a point (X) is reached where the vegetation crashes; the system becomes unstable because of overgrazing. As the herbivore population drops off sharply, the vegetation regrows until a second point (Y) is reached; the system becomes unstable again and a vegetation bloom occurs. The vegetation crash and bloom are **catastrophes** in the mathematical sense of the word. X and Y are known as **critical thresholds**. The spruce budworm population shows this type of behaviour (Ludwig and others, 1978), with the herbivore following the crash and outbreak pattern (see Figure A-1b). This type of catastrophe is called a fold (see Figure A-1c). For P_1, P_3 there is one value of X (X_1, X_3) but for P_2 there are two possible values, X_2 and X_4. Which value the system takes at P_2 depends on its history (i.e., depends on the path that the system is following).

A more complicated form of catastrophe is shown in Figure A-2. This is the Riemann-Hugoniot catastrophe. This behaviour is more complicated than the fold because the system may move along an equilibrium path with a catastrophe threshold (A to B) or one without (C to D). It may also exhibit bifurcations or divergences or multiple paths. The Riemann-Hugoniot catastrophe is still quite simple, being of order 3. There are examples up to order 7.

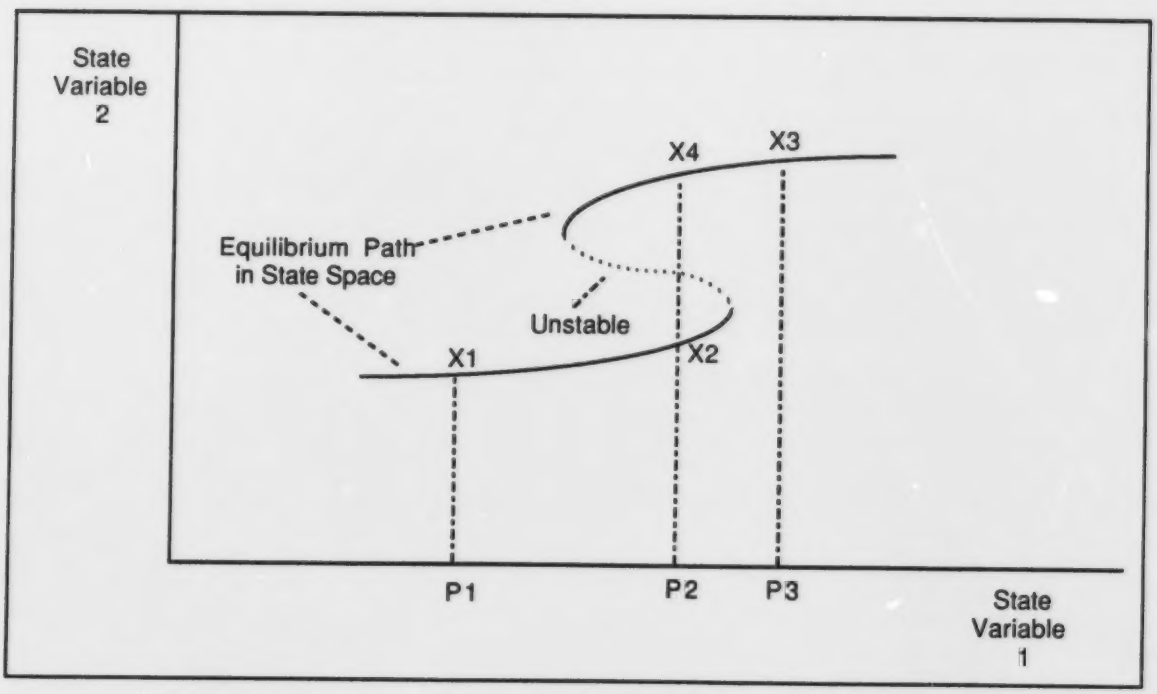


Figure A-1c: The basic fold catastrophe. There is an equilibrium path through state space. The solid curves represent the stable part of the path. The dotted curve represents the unstable part of the path. (The system cannot stay on the unstable part, only on the stable part.) For points P1 and P3, there are single values of X possible. The value of X at P2 depends on which curve the system is currently following. The catastrophe is the system jumping from the upper to the lower curve, or vice versa. This occurs at the thresholds (X and Y in Figure A-1a).

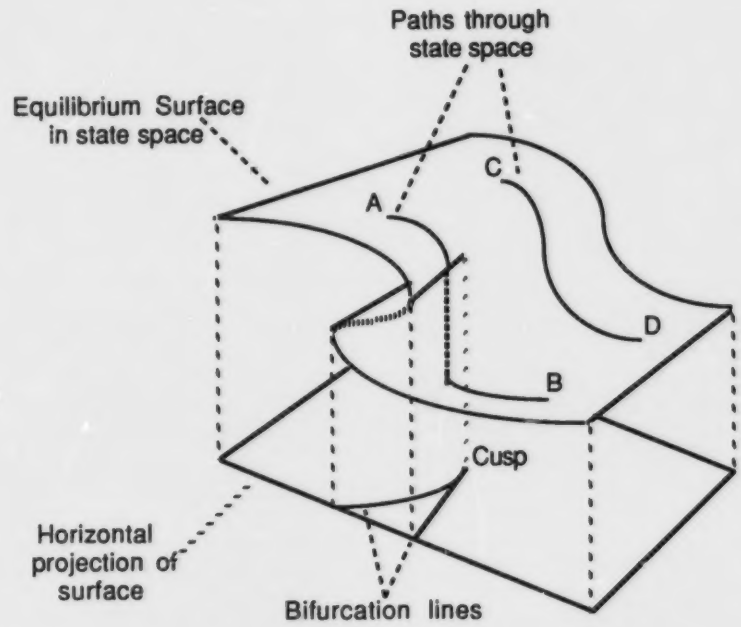
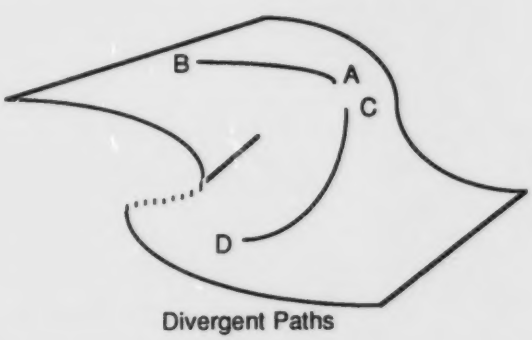
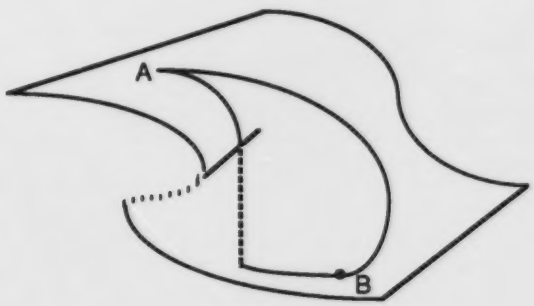


Figure A-2: The Riemann-Hugoniot Catastrophe. The set of equilibrium points for the system is a surface in state space. The path followed by the system may pass through a catastrophe threshold (from A to B above, the dotted part of the curve corresponds to the jump) or it might not (from C to D). The system may hit a cusp (see above), which is a bifurcation point. (The system may follow either bifurcation line after the cusp; which one it follows is not predictable.) Two points in state space (A and C below left) may be very close, but the system may diverge quite dramatically later on in its development (from A to B versus C to D). Also, there may be different paths that the system can follow from A to B (see below right). One path may involve a catastrophic system change (the dotted line), while another does not.



Divergent Paths



Multiple Paths

The point is that systems can exhibit very complicated and dramatic behaviour, even when they are deterministic. The lesson is, be prepared for surprise. To explore these notions in more detail, see Holling (1986) or Jones (1975) for excellent discussions.

An observant reader will note that the notion of equilibrium used throughout this appendix is different than the one used in the main body. Equilibrium in a stability sense is different from equilibrium in a thermodynamic sense. The former refers to a balance of the forces acting on a system and has its origin in classical Newtonian mechanics. **Thermodynamic equilibrium** refers to a system state which is uniform throughout and indistinguishable from its surroundings. For a biological system, this represents death. Non-equilibrium thermodynamics is about the organization of systems such that they are not in thermodynamic equilibrium. Needless to say, this difference causes great confusion. Non-equilibrium systems can be stable, that is, the forces acting on them are in equilibrium!

Organization refers to changes in the function of a system and its internal connections (structure), so as to better carry out some organizational imperative. The end point in the process of organization represents an optimal trade-off between the different objectives of the system. It is in this sense that the phrase **optimum operating point** is used. For more detailed discussion of the organizational imperatives of ecosystems and the trade-offs involved, see Kay (1984, 1989).

Thermodynamic branch is used to indicate the path through state space followed by an ecosystem as it develops from (thermodynamic) equilibrium to its optimum operating point (i.e., steady-state stable equilibrium point in state space, which is far from thermodynamic equilibrium) under normal conditions. An example would be the path normally followed by an ecosystem as it develops from a bulldozed field (thermodynamic equilibrium) to a climax forested ecosystem (optimum operating point).

Appendix B: Ecological Stability

The following is meant to give the reader a taste of the varying definitions in the literature related to the term **ecological stability**. The discussion is detailed but by no means complete.

When an ecosystem is described as being stable it usually means that it is, in some sense, well behaved. Many attempts have been made to formalize this definition using mathematics. The most natural approach is to use a definition of stability commonly used in the physical sciences, that is Lyapunov stability. This requires that some function be found that describes the system and that satisfies Lyapunov's stability criteria (see Lewontin (1969), and Harte and Levy (1975), for a review of stability theory as applied to ecology).

Many workers have attempted to use the stability of the species population to define ecological stability. Usher and Williamson (1976) state: "Roughly speaking, ecological stability is the strength of the tendency for a population or set of populations to come to an equilibrium point or to a limit cycle, and also, related to that, the ability of a population system to counteract disturbances." Many articles and books have been written using this definition of ecological stability (for the ultimate see Granero-Porati and others (1982)), but an equal number of papers and books have been written challenging this approach.

Hirata and Fukao (1977) and others have used the biomass of species as the important function that must be stable. This is a slightly more flexible approach because it allows for fluctuations in populations as long as the total biomass is stable. Others have talked about the stability of the functioning of a species, that is, their niche remains stable. More recently, Bormann and Likens (1979) have discussed stability in terms of the functioning of the entire ecosystem, as measured by stream water input. Another suggestion is that the stability of the structure (i.e., food web) of the ecosystem characterizes ecosystem stability. Presumably, this would be measured using the measures developed by Rutledge (1976) and Ulanowicz (1979, 1980, 1986). Still others have suggested that the stability of the macro- (i.e., external) or micro-environment are the important characteristics. (This would be measured by such things as temperature fluctuations, rainfall fluctuations, and humidity fluctuations.) Unfortunately, no one of these system measures is sufficient to characterize ecological stability. Any one of them may not be stable, in a Lyapunov sense, while the system as a whole may be 'well behaved'.

In the last few years, the phrase ecological stability has been used to mean the stability of so many different characteristics of ecosystems that *one must be very careful to understand what authors mean when they use it.* Clearly such confusion is undesirable. Several authors have suggested that a broadening of the definition of stability is necessary if it is to be usable in an ecological context. What follows is a sampling of the ideas of a few authors.

Preston (1969) states: "Stability lies in the ability to bounce back. An ecological system may be said to be stable, from my point of view, during that period of time when no species becomes extinct (thereby creating a vacant "niche") and none reaches plague proportions, except momentarily, thereby destroying the niches of other species and causing them to become extinct." This is an interesting definition because it does not require that the populations be stable in the Lyapunov sense, only that they be non-zero.

Rutledge (1974) identifies three different properties of ecosystems, all of which should be encompassed under ecological stability. The first is the sensitivity of the components of the ecosystem to perturbation. The larger the sensitivity, the less the stability. The second is the persistence of the ecosystem over time. The longer it has survived, the more stable it is. The final property is the ability of the ecosystem to return to its equilibrium state after being perturbed from it.

May (1974) identified three tributaries to the stream of ecological stability theory: "One draws inspiration and analogies from thermodynamics, and is concerned with broad patterns of energy flow through food webs. A second theme, . . . deals with the physical environment, and the way it limits species' distributions and affects community organization. A third tributary concentrates on the way biotic interactions between and within populations act as forces moulding community structure."

Margalef (1975) is a little more pessimistic and suggests that "it is perhaps questionable whether the term stability should be retained, as it has been used too much in different and divergent speculation." Wu (1974) suggests that it is more relevant to talk about ecosystem health, where an ecosystem is considered healthy if its state variables are within a certain range.

My own opinion is that the idea of stability should be kept, but only in the narrow confines of the Lyapunov definition. To define what is meant by a well-behaved ecosystem, other ecosystem properties, besides stability, must be defined and quantified. A number of authors have attempted this. Many of them work with the idea of a dimensional state space. Usually each *n* axis corresponds to the population of one of the species. However, other state variables can be used as well. There are a number of points in this hyperspace that are stable equilibria of the ecosystem. About each of these stable points is a cloud. If the system is displaced from an equilibrium, but remains within the cloud, it will return to the initial equilibrium point. If it is displaced outside this cloud, it will move to some new stable equilibrium state.

Holling (1973) introduced the idea of resilience. He defines resilience as the minimum distance from the equilibrium point to the edge of the cloud. Thus, resilience is measured by the minimum disturbance necessary to disrupt the system and cause it to move to a new equilibrium state. Stability is the degree of oscillation the system exhibits about its stable equilibrium point. Holling points out that forests that undergo pest outbreaks, such as the spruce budworm, are unstable. They experience extreme oscillations in populations. Yet the system almost always bounces back to its original state. It is resilient. Holling notes that resilient systems normally are not stable and vice versa. Hill (1975) expands on Holling's idea and observes that there are two kinds of stability involved. One is 'no-oscillation stability' and refers to the stability of the state variables in the absence of stress. The other he calls 'stability-resilience'. This refers to the stability of the state variables while the system is under stress and after the stress is removed. This latter stability refers to the degree of oscillation (flutter) the system experiences while under stress, and how quickly this is dampened when the stress is removed.

Cairns and Dickson (1977) have examined the stability-resilience of stream ecosystems. They have identified four properties of ecosystems that determine the stress recovery characteristics of ecosystems: vulnerability, elasticity, inertia, and resilience.

VULNERABILITY: the lack of ability to resist irreversible damage (which is defined as damage that requires a recovery time greater than a human life span). Presumably, it is measured by the size of disturbance necessary to cause irreversible damage.

ELASTICITY: the ability to recover after displacement of structure or function, to a steady state closely approximating the original. Presumably, this is measured by the rate of recovery after disturbance.

INERTIA: the ability of an ecosystem to resist displacement or disequilibrium in regard to either structure or function. Presumably, it is measured by the size of the disturbance needed to displace the system.

RESILIENCE is the number of times a system can undergo the same disturbance and still recover.

Cairns and Dickson are not clear about how to measure these properties or about the difference between them. But, they do point out that the size of the disturbance necessary to displace the system, how far the system can be displaced before it will not bounce back, how long it takes to bounce back from the disturbance, and how many disturbances the system can tolerate, are all properties that influence the reactions of an ecosystem to stress and need to be understood in detail.

Orians (1975) has identified seven properties of ecosystems that are related to their stability:

CONSTANCY: the lack of change in some parameter of the system.

PERSISTENCE: the survival time of the system.

INERTIA: the ability to resist external perturbations.

ELASTICITY: the rate at which the system returns to its former state following a perturbation.

AMPLITUDE: the area over which the system is stable (the same as Holling's resilience).

CYCLICAL STABILITY: the property of a system to cycle about some central point or zone.

TRAJECTORY STABILITY: the property of a system to move toward some final end point or zone despite differences in the starting points.

The factors that increase each of these properties are listed in Table B-1.

Table B-1: Environmental Factors and Phenotypic Characteristics of Species that Increase Different Kinds of Stability (after Orians 1975)

A. Persistence

1. Environmental heterogeneity in space and time.
2. Large patch sizes.
3. Constant physical environment.
4. High resource-utilization thresholds of predators.

B. Inertia

1. Environmental heterogeneity in space and time.
2. Greater phenotypic diversity of prey.
3. Multiplicity of energy pathways.
4. Intra-specific variability of prey.
5. High mean longevity of individuals of component species.

C. Elasticity

1. High density-dependence in birth rates.
2. Short life cycles of component species.
3. Capacity for high dispersal.
4. Strong migratory tendencies.
5. Generalized foraging patterns.

D. Amplitude

1. Weak density-dependence in birth rates.
2. Intra-specific variability of component species.
3. Capacity for long-distance dispersal.
4. Broad physical tolerances.
5. Generalized harvesting capabilities.
6. Defence against predators not dependent on a narrow range of hiding places.

E. Cyclic Stability

1. High resource-utilization thresholds.
2. Long lag times in species responding to changes in resource availability.
3. Heterogeneity of environment in space and time.

F. Trajectory Stability

1. Strong organism-induced modifications of the physical environment.
2. All factors increasing elasticity.

Orians believes that an understanding of these properties can be obtained only from an understanding of the interactions of species and an appreciation of the past disturbances and selection pressures that have acted on the species. We must examine stability from this perspective using a precise definition of the property of an ecosystem we are trying to understand, and in the context of a specific type of disturbance.

Robinson and Valentine (1979) review the idea of stability and introduce their version of the concepts of elasticity, invulnerability, and invadability. Van Voris, O'Neill, Emanuel, and Shugart (1980) introduce the notion of functional stability.

More recently DeAngelis and others (1989), in the context of nutrient dynamics and food web stability, identify five types of stability. These are:

LOCAL STABILITY: the tendency of all system components to return to their steady-state equilibrium values following small perturbations.

STABILITY TO LARGE PERTURBATIONS: the possibility that the system flips to a new steady-state equilibrium point when subjected to large perturbations.

RESILIENCE: the rate of return of the system to its steady-state equilibrium point after a disturbance.

STRUCTURAL STABILITY: the possibility that small gradual changes result in a catastrophe.

PERSISTENCE: the tendency for the components of a system to stay within specific positive bounds through time.

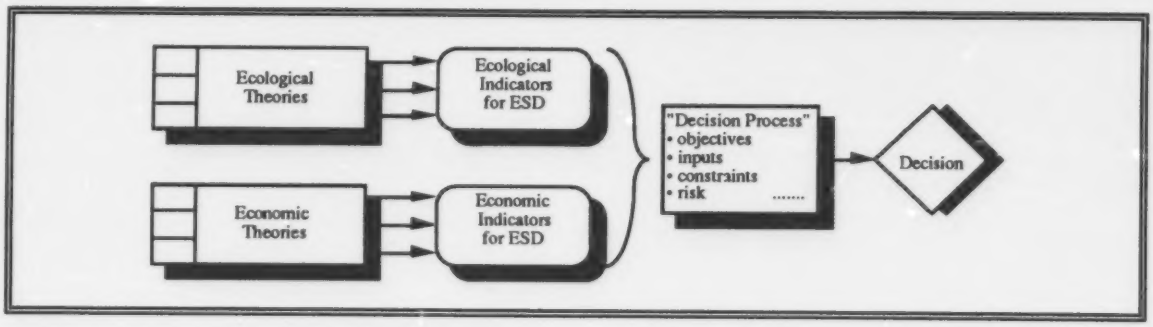
DeAngelis and colleagues end their paper by noting: "In fact, even the existence of steady-state equilibria in natural ecosystems, on which four of the five types of stability considered here depend, is open to some question. Nevertheless, it is doubtful that ecologists will abandon stability as a concept in the near future, so critical assessments will continue to be of value."

Holling and his colleagues have introduced the use of catastrophe theory in ecological systems (Ludwig and others 1978; Jones 1975; May 1977; Holling 1986). The last of these references is an excellent, readable overview of Holling's ideas about dynamic stability and surprise, and the reader is encouraged to pursue it.

Clearly, before any real understanding of ecosystem response to environmental change can be obtained, the confusion about concepts that fall under "stability" or "well-being" must be dealt with.

The Role of Indicators in the Decision Process

H. Jack Ruitenbeek



Abstract

Whether indicators of sustainable development are derived from ecological theory or economic theory, the way they are incorporated into the decision process can influence policy prescriptions and, in some instances, indicator design. The decision process – at the individual and collective state levels alike – further complicates the design and choice of indicators. To highlight this, a generic decision process is discussed using indicators as objectives, inputs, or constraints. In addition, risk and uncertainty are identified as important components of decisions about sustainable development. Some analytical techniques – and the requirements they impose on indicators – are reviewed .

Contents

1 Background 60

2 The Decision Process: Some Issues in Political and Welfare Economics 61

3 Indicators as Objectives, Inputs, and Constraints 67

4 Indicators, Expectations, and Risk 73

5 Concluding Remarks 79

References 80

Appendix A: Arrow's Axioms Required of Collective Choice Processes 84

Appendix B: Attributes of Selected Approaches to Economic Analysis 85

Appendix C: Attributes of Risk Analysis Mechanisms 90

RUITENBEEK, H.J. 1991. The Role of Indicators in the Decision Process. In: CANADIAN ENVIRONMENTAL ADVISORY COUNCIL. *Economic, Ecological, and Decision Theories: Indicators of Ecologically Sustainable Development:* Ottawa: Canadian Environmental Advisory Council (CEAC). Prepared under contract for the CEAC Workshop on Indicators for Ecologically Sustainable Development (Ottawa, July 1990).

1 Background

As we are making decisions every day, we often take the decision process somewhat for granted and, in the extreme, disavow that it is even a factor in shaping results. This paper, therefore, attempts to provide a discussion framework that allows us to step back and look more critically at the decision process as another step between gathering information and turning that information into actions that will promote ecologically sustainable development (ESD). Although many of the issues raised in this paper will be familiar, it is hoped that they will provoke enough interest to stimulate further focused discussion of these issues in the context of designing ESD indicators.

The idea that ESD is a desirable economic policy goal is being more widely professed within both the ecological and economic disciplines. Macro-economic policy-makers are facing the often difficult challenge of designing policies, programs, or projects to promote this goal. At the same time policy analysts are faced with the challenge of selecting and analysing the most effective means for measuring policy effectiveness. Keen interest has thus arisen in developing indicators that somehow reflect the principle of ESD.

The role of indicators in ESD is driven by both the theoretical paradigms (ecological or economic) that underlie them, and their eventual role in the decision process. The two companion papers by Kay and Victor in the present volume address some of the linkages between the various ecological and economic theories or approaches, and the indicators that can be derived from these approaches. A key conclusion of these papers is that the appropriate set of ESD indicators will in itself depend on the theoretical paradigms from which they are derived; there is no general theoretical consensus on what an appropriate indicator of ESD is.

Even if such indicators are generated, however, it is clear that they must eventually be translated into a final decision and accompanying action. The steps and the analytical techniques that are applied at this stage can be independent of those used in generating the indicators in the first place. It could be just as important to describe how a given indicator is used within the decision process as it is to describe its initial derivation or theoretical foundation. But the connections between ESD indicators and the decision process are often not explicitly considered in discussions of sustainable development.

The general objective of this paper is to concentrate on issues that are not specific to a particular ecological or economic theory, but that are nonetheless relevant in drawing connections between ESD indicators and the decision process.

As shown conceptually in Figure 1, any potential indicator of ESD – whether a measure of environmental integrity, human welfare, or economic wealth – will eventually be treated as a specific unit of information ('I'), which is ultimately translated into some action ('A'). The actual decision process ('D') can be thought of as an analytical framework, which itself serves to operationalize the myriad and sometimes contradictory indicators. The scope of this paper involves addressing issues related to the decision process that are believed, *prima facie*, to be relevant to the issue of sustainable development. In particular, the paper will consider the following:

a) *The decision process* – There is no collective decision process that will generally achieve an optimal allocation of resources. We discuss the importance of describing the collective and individual decision processes that do exist, as different descriptions of the decision process could require different types of indicators. (Section 2)

b) *Roles of indicators as objectives, inputs, or constraints* – Any given indicator can, in the decision process, generally take on one of three roles. We discuss the importance of identifying what role a given indicator is meant to play within a given decision and will show, through a simple example, how different treatment of indicators might translate into different policy decisions. (Section 3)

c) *Role of indicators in addressing expectations, risk, and uncertainty* – Risk and uncertainty are often important attributes of a decision process and various analytical techniques are available to address these factors. We shall briefly delineate the various techniques and outline the special requirements that such techniques have on identifying and using ESD indicators. (Section 4)

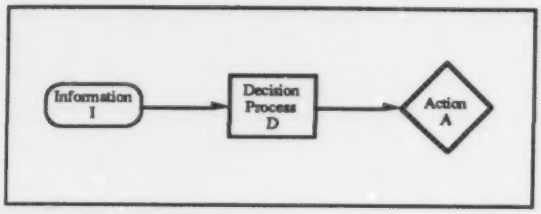


Figure 1: Decision

The reader is forewarned that the approach taken in this paper relies very much on aspects of decision theory addressed in the economics literature. A vast body of literature also exists in other disciplines – philosophy, political science, psychology, sociology, and business administration, to name but a few. We concentrate on economics literature because many decision-makers rely at least partially on economic rationales for their actions.

Although the discussion draws on many aspects of economics literature, it relies more precisely on work covered in welfare economics and political economics. In decision theory, welfare economics has traditionally concentrated on how individual 'agents' make choices. It investigates how any single agent – a consumer, a firm, or even government as a whole – makes decisions when presented various choices. Concepts of economic allocative efficiency and distributive equity play an important role in the principles of welfare economics. Political economics, again as applied to decision theory, has complemented this by concentrating on the existence and nature of *collective* choice mechanisms, how these mechanisms are reconciled to individual needs, and how they incorporate other political issues such as ethics and civil liberties. The literature thus lets us consider decisions in a generalized framework that can address matters of efficiency, equity, ethics, and uncertainty, all of which play important roles in discussions of ESD.

This literature is also of interest because, to a large extent, it has formed part of an active debate over whether economic structures should be market-oriented or coercive through government regulatory intervention. Interestingly, arguments promoting sustainable development often differ on which is more desirable: market solutions or government coercion. Although it is beyond the scope of this paper to address these arguments, the issues raised here will allow us to revisit some of the theoretical underpinnings of this debate and, in the process, shed some light on the role of both the collective and individual decision process in achieving sustainable development.

In addition, it is recognized that the nature of the decision-maker – consumer, administrator, policy-maker, manager, scientist – can itself impose quite different information demands on indicators. The idea that indicators must be designed for the appropriate target audience is generally acknowledged as a vital consideration in formulating and reporting indicators. Rather than distinguishing between different target audiences, however, this paper will focus on concepts that are not necessarily audience-specific. Where a specific audience for indicators is implied, we shall generally focus on policy-makers or individual members of the public who influence policy-makers.

Some of the basic concepts discussed here will be familiar to those working in applied or theoretical economics. Economics literature on decision theory does not, however, explicitly address issues relating to ESD indicators. This paper thus complements this literature by presenting the issues in a manner that will be helpful to policy-makers trying to come to grips with how these indicators might eventually be incorporated within their decisions.

This paper does not advocate the use of any particular decision technique or process; however, at the end of each section it presents suggestions for where Canada might direct some of its efforts in designing types of indicators.

2 The Decision Process: Some Issues in Political and Welfare Economics

2.1 Introduction: Why Describe the Decision Process?

It is natural to ask, first, why we should concern ourselves with how decisions are made. There is a natural tendency to assume that if there is enough technical and economic information before us, then somehow we should be able to translate that information into appropriate and correct collective actions. In such a world, the only relevant task for the ecological and economic sciences is to deliver to decision-makers a host of indicators that can be used to guide decisions. The collective decision will then follow naturally and deterministically from whatever information is provided.

Further, as noted in the Kay and Victor papers, there is still considerable theoretical debate over which indicators are relevant for monitoring sustainable development from either an ecological or an economic point of view. Given the lack of consensus within the disciplines, we might well be reticent about extending the debate further still by questioning the integrity of the decision process itself.

Nonetheless, a central position of this paper is that even if we have no quarrel with the theoretical basis or measurement of a given set of indicators, the manner in which this information is used within the decision process is in itself significant. The final outcome of policy deliberations and, implicitly, the route to sustainable development, will depend on both the nature of the decision mechanism and the role of indicators within that mechanism.

In particular, the questions that we shall pursue in this section relate to the existence and nature of collective choice mechanisms:

a) Assuming that all the relevant indicators are available, does an acceptable 'optimal collective decision process' (D_O) exist for society, which will translate these indicators into a 'correct' action?

b) If D_O does not exist in practice, to what extent is the decision process (D) consistent with D_O, how has D itself evolved, given that it is not the optimal decision structure, and especially, are the indicators we require dependent on the type of decision process that exists?

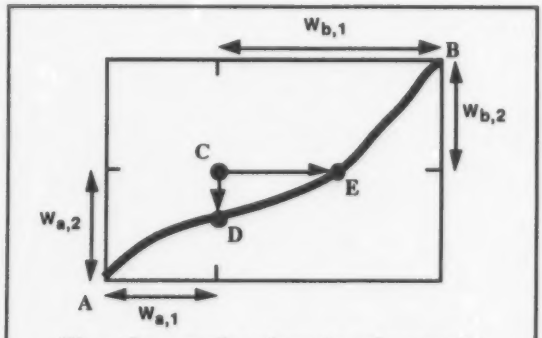
The short answer to these questions is that D_O – the optimal collective decision process – does not exist. Existing sub-optimal processes represent different decision structures, which in turn require different types of indicators.

We shall elaborate on each of these points, within the context of particular sustainable development issues. Section 3 addresses in more detail the role of indicators within any given decision process.

2.2 Markets and Governments: a First Look at Different Decision Structures

Similar to the ecological literature that deals with optimal operating points for ecological systems, economic choice theory is replete with discussions of whether there does, indeed, exist some 'optimal' operating point for society as a whole. The role of a decision process is, presumably, both to identify that point and then to identify some means of getting there. Before we consider the central question of whether such a decision process exists, it is useful to review briefly some of the underlying rationale for government involvement. For convenience, we shall cast the discussion in terms of the conventional 'market' versus 'government' nomenclature.

Proponents of the market approach often argue that the market is an efficient mechanism for allocating society's scarce resources. This result is, in fact, part of the well-known first theorem of welfare economics. What they often fail to add, however, is that there is more than one potential point of equilibrium and that some equilibrium points may be socially preferred over others for distributive reasons. The usual presentation of this result involves a concept known as the Edgeworth Box, shown in Figure 2, for two goods and two individuals. The locus joining points A and B represents a set of competitive equilibria. If a system starts at point C, off the locus, a competitive market will eventually re-allocate the resources to some point on the equilibrium locus (between D and E in this case). Once on the path, the market will *not* re-allocate goods to a different point. In theory, for example, a point on the locus very close to A (where individual B gets most of the goods) is defined as efficient as it is on the locus.



$W_{a,1}$: Consumption of good x_1 of person A.
$W_{a,2}$: Consumption of good x_2 of person A.
$W_{b,1}$: Consumption of good x_1 of person B.
$W_{b,2}$: Consumption of good x_2 of person B.

Figure 2: Edgeworth's Box. From C→D or C→E? Both are efficient equilibria...but which is best?

Within this framework, there are two roles that political economists often ascribe to government. The first is a role of reform – under some circumstances market failures prevent market forces from achieving even allocative efficiency. The most common failures are externalities such as public goods or the existence of untraded commodities such as pollution. A reform government has a role in correcting these failures. In our Edgeworth Box, a market failure might cause the economy to be stuck at point C – government intervention nudges the economy toward efficient production.[1]

A second potential role is that of a distributive government. This involves moving the economy from one point of equilibrium, say D, to another point, say E, if some social choice process determines that E is better than D for some reason – equity, for example. The most common mechanism for achieving this is, of course, taxation.

Even with this simple distinction, we see that in both these cases – reform and distribution – the planner or decision-making authority is asked to fulfil a function that complements that of the market. Intuitively, one would expect that the planner's information requirements normally become more complex with this added responsibility.

A simple thought experiment to demonstrate this point involves tradeable emission rights. A common reformist prescription for correcting the failure of markets to allocate environmental goods is to create markets for these goods by attaching property rights to environmental services. (See, for example, Block 1990 or Hahn and Hester 1989.) One example of this involves tradeable emission rights.

Suppose that we require that any industry choosing to pollute must be in possession of an adequate quantity of emission rights. Suppose further that some ecological or economic theory tells us that the optimal level of sulphur emissions in a given air shed is 1 million tonnes per year. We know that if we place 1 million t/a of emission rights on the market, an optimal amount of pollution will occur and it will be achieved at least cost: an efficient solution is achieved. But how should these rights be allocated? Should they all be given to industries in the region and then be allowed to trade? Should they be auctioned off to the highest bidder? Should they be given to a local Fish and Wildlife Association that can then turn around and sell them to the industries that need them? In each case, we know that the resulting allocation after trading will be efficient, but there will be different distributions of wealth in each case.

If the central decision-makers are interested only in efficiency, then all they need to know is the optimal level of pollution. If they are also interested in, say, equity or usufructory property rights,[2] then they may require more information on the preferences of individuals before deciding how to allocate emission rights.

Even in this relatively simple example, we see that the two decision structures – market economy and government planner – appear to have different information requirements.

2.3 Markets and Governments: Is the Right Answer Achievable?

To this point we have essentially assumed that a decision process existed that was capable of identifying and achieving some optimal operating point for society. We saw that market structures might not achieve that point in the presence of market failures, and that there might therefore be some role for collective decisions. Within the discipline of political economics, however, there has been considerable discussion over whether a social choice mechanism does indeed exist that is capable of achieving a given set of economic, social, and political goals. If such a mechanism existed, one might then argue that it is not so critical to look at the information requirements of different decision mechanisms.

The straightforward answer to the existence question is no, there is no single decision process capable of giving us the correct answer even if all the information is available. This result is attributed originally to Arrow (1963), and revolves around a possibility theorem that proved there was no collective decision process

[1] In its minimal form, this will involve enforcement of property rights; market processes themselves will not do this. Technically, such enforcement is a public good and will be undersupplied by private agents.

[2] We shall return to this point in Section 3. It essentially argues that common ecological resources, while having had no formal ownership rights attached to them, do in fact belong to whomever has been using them. Industrial polluters, on the one hand, can use this position to argue that they should be allocated the rights as they have been actively using the resource, but it seems more compelling to use the same argument for allocating the rights to anyone who has been passively using the resource.

that concurrently satisfied five apparently plausible requirements.

Appendix A outlines Arrow's axioms, but they essentially require that any collective decision process:

P: be capable of identifying Pareto-efficient allocations of resources to correct market failures;

ND: be non-dictatorial;

U: have an unrestricted domain to the extent that it could choose between any proffered alternatives;

R: be rational; and,

I: be independent of irrelevant alternatives.

Each of these conditions appears relatively innocuous, but Arrow demonstrated in his famous Possibility Theorem that these five conditions can not be simultaneously satisfied by a collective choice process. The nature of the proof was such that any choice mechanism that satisfied conditions P, U, R, and I would then need to be dictatorial. As such an arrangement is not generally acceptable, we are left with dropping one of the other options to find a social choice function.

Some systems, those advocating egalitarian distribution for example, will often implicitly give up P. In such cases, the mechanism will fail to find efficient allocations of resources. In addition, however, Wilson (1972) showed more generally that, without P, the resulting mechanism is always indifferent between alternatives regardless of individual preference ranking. Under such circumstances, the social choice mechanism would not need to give any consideration to individual preferences.

The conditions most commonly dropped in search of a collective choice mechanism are U, R, or I, but in all such cases administrative efficiency must be sacrificed. Dropping U limits the effectiveness of the system to make choices. Dropping R requires that more information be gathered before a decision can be made. For example, in a rational system, if indicators tell us that society prefers clean water to clean air and clean air to pretty landscapes, we safely conclude that clean water is preferred to pretty landscapes. Dropping the rationality restriction would require that we find out if the last conclusion is indeed true.[3] Finally, dropping I leads to the greatest amount of administrative inefficiency in that we must then get information on all feasible consumption choices, even if they appear to have nothing to do with the decision at hand. In choosing between clean air and clean water, for example, we must know how people feel about pretty landscapes.

It is thus generally concluded that decision processes must sacrifice either allocative efficiency, decision-making efficiency, or democracy.

The major inference of this literature is that market structures, government structures, or hybrids are not necessarily inherently capable of providing an ideal decision process. However, an actual decision process is typically some second-best hybrid.

From a practical perspective, this implies that we cannot presume that one particular decision model is most appropriate for designing indicators. Different decision structures may have different information requirements.

2.4 Markets, Democracies, and Dictators

As a first step, it is useful to review briefly some of the decision-making mechanisms that actually exist, with a view to identifying whether different information requirements might exist for different types of structures.

To indicate the potential information requirements, Figure 3 shows, in order of increasing complexity, six potential sets of environmental information. Figure 4 then summarizes some of the major types of social choice mechanisms in the literature. In each case we provide our own brief assessment of the role of environmental information – and implicitly of ESD indicators – in each of the decision mechanisms. Some interesting generalizations can be made from this review.

First, non-coercive decision structures such as competitive markets (CM) or voluntary exchange (VE) mechanisms can be expected to generate some of their own information requirements. But their disadvantage is that many of the global

[3] Those familiar with political and individual decision processes often argue that decisions are not made rationally; we should then concentrate on designing indicators that target irrational approaches.

indicators that might be relevant in a coercive structure will not be gathered in CM or VE structures. As they become public goods, no incentives exist for individual agents to spend their resources collecting or generating such indicators, since information cannot, generally, be sold.

x:	all privately accessible technical information
y':	all privately accessible and some public technical information
y:	all publicly accessible relevant technical information
z:	all publicly accessible (including irrelevant) technical information
	$(z \supseteq y \supseteq y' \supseteq x)$
Py:	all individual preference information on y
Pz:	all individual preference information on z

Figure 3: Environmental Information Sets

Second, the nature of coercive dictatorial processes, even under benevolent dictators, tends to be information intensive. In effect, the dictator needs to know technical information about environmental goods and services, as well as people's preference orderings for these goods and services. Further, a purely benevolent dictator (BD) with no access to incentive mechanisms for preference revelation, such as those available in the MDP or DR processes, will require complete information about individuals' preferences on items that may not be relevant to the immediate allocation decision.

Third, democratic processes are generally more information intensive than non-coercive processes, but they require less information than dictatorial processes. In particular, systems of preference revelation available within democratic processes are not available within dictatorial processes. An exception to this involves the preference intensity processes (PIP), which tend to be very information-intensive decision mechanisms.

A final observation relates to the set of mechanisms referred to as structurally induced equilibrium (SE) processes. Conceptually, such processes produce an allocatively efficient outcome by imposing an external institutional structure that prevents voting cycles (i.e., Axiom I is retained).[4] These structures also have shortcomings; rationality (R) is sacrificed and they may be subject to other types of manipulation such as agenda setting. But an interesting conclusion relating to SE mechanisms is that they suggest that the decision process itself might evolve structurally as a result of the type of information made available in society.

We are perhaps by now getting used to thinking that decision structures will tend to determine what sort of indicators may be required. But this last observation suggests that the very types of decision structures that evolve might be influenced by the type of information that becomes available. If we gather just technical information of a purely public nature and provide it to free market participants, will this tend to promote competitive and voluntary exchange types of processes? If we gather information on individuals' preferences for environmental goods, will this lead to more opportunities for government intervention? Although such questions are just starting to come onto the research agenda and cannot be answered at this time, policy-makers may wish to reflect on them in selecting types of information gathering and dissemination processes to be pursued.

2.5 Summary: Implications for Indicators

This section did not so much challenge any existing decision process, as illustrate that the decision process is itself a variable that must be considered in designing or selecting ESD indicators.

A simple summary of the effectiveness of eight collective decision structures – ranging from no government to an active reform and distributive government – is shown in Figure 5. Although the structures and their effectiveness are based

[4] The impetus behind pursuing such mechanisms was primarily that many other processes were subject to a phenomenon known as strategic voting or vote manipulation (Farquharson 1969). Gibbard (1973) and Satterthwaite (1975) proved generally, by adding a sixth axiom to Arrow's set – that of positive association (PA) – that democratic processes considering more than two alternatives were always open to vote manipulation. Hence SE mechanisms sought to impose structures that set voting agendas or other structures, which got around this problem. Romer and Rosenthal (1979; 1982) have found in empirical work, however, that such structures suffer from another type of inefficiency imposed by 'agenda setters' who, in a sense, begin to undermine the democratic nature of the structure.

Decision Mechanism	Attributes of Mechanism	Role of Environmental Information as a Publicly Supplied Good - Planner's Requirements -
Non-coercive Competitive Market (CM)	• No government intervention. • Utility maximizing private agents. • Failures arise with public goods, externalities, increasing returns, asymmetric information.	*None (x)* Individual agents will gather technical information of relevance to them, but only if the information can be kept private.
Non-coercive Voluntary Exchange (VE) Wicksell 1896 Lindahl 1919 Johansen 1963	• No government intervention. • Formation of collectives that maximize utility subject to unanimity within collective. • Failures arise as each individual has an incentive to appear to be the least advantaged in the group.	*None (y')* Some public information is gathered by individuals in collectives, but at a sub-optimal level.
Coercive Benevolent Dictator (BD) Heal 1973	• Allocation by single authority. • Maximizes some form of social welfare. • Failures arise as individual citizens have incentive not to reveal preferences to authority.	*Extreme (Pz)* Information required by planner on individual preference orderings for all potential environmental goods and services.
Coercive MDP Mechanism (MDP) (attributed to: Malinvaud 1971, Drèze & de la Vallée Poussin 1971)	• Allocation by single authority. • Maximizes some form of social welfare. • Individual citizens provided incentives to reveal their preferences to authority. • Failures arise as cooperation is not always dominant strategy.	*Significant (Py)* Information required by planner on individual preference orderings for environmental goods and services most likely to be preferred.
Coercive Demand Revealing (DR) Vickrey 1960	• Allocation by single authority. • Maximizes some form of social welfare. • Individual agents provided incentives to cooperate with preference revelation through tax structure. • Cooperation is dominant strategy for individuals. • Failures arise as cooperation is not dominant strategy for collectives or coalitions.	*Significant (Py)* Information required by planner on individual preference orderings for environmental goods and services most likely to be preferred.
Democratic Majority Rule Processes (MRP) May 1952	• Allocation by majority vote. • Failures arise because of voting cycles (cyclicity or intransitivity); existence of equilibria not always obvious.	*Moderate (y)* Information required on those environmental goods and services most likely to be preferred, but not people's preferences of one over the other.
Democratic Preference Intensity Processes (PIP) Mueller 1979 Inman 1987	• Allocation by weighted vote (e.g., point voting, vote trading). • Drop Arrow's axiom I: Independence of Irrelevant Alternatives. • Failures arise as voting behaviour can lead to inefficient allocations.	*Significant (z)* Information required on all environmental goods and services, i.e., not just those likely to be preferred.
Other Structurally Induced Equilibrium (e.g., MRP-SE) Plott 1967 Shepsle 1979	• Allocation by majority vote. • Choice is determined by structure of decision-making group (e.g., committee, jurisdiction, amendments). • Failures arise out of potential oligarchical behaviour leading to inefficient allocations.	*Moderate (y)* Information required on those environmental goods and services most likely to be preferred. Information that is made available could itself change the structure of the decision process.

Figure 4: Summary of Examples of Collective Decision Mechanisms

on Inman (1987), the additional entries relating to the indicator requirements are based on the above discussion. It may seem excessive to include categories such as anarchy and benevolent dictatorships within such a summary, but society is a medley of different decision points. Decisions in some areas are subject to very little outside intervention, whereas others, such as spending within a single company or government department once the budgets have been set, may be essentially dictatorial.

Some general conclusions relating to the relationship between indicators and decision processes are:

a) There is no single deterministic optimal collective decision process; indicators must, therefore, be sensitive to the different requirements of different decision processes.

b) Practical structures typically involve some combinations of market structure and government coercive action; these generally have different requirements for ESD indicators.

c) A given decision mechanism might itself be a product of the type of information that is made available; this suggests that we might, in effect, manipulate the type of decision processes that evolve by the types of information we make available.

Finally, there are two specific questions we have not considered here that arise naturally out of the above discussion. First, should we have explicit indicators of how much of a particular environmental commodity falls within markets and outside of markets? Second, is one decision-making mechanism more appropriate than others in addressing issues of sustainable development and, if so, should we reform or augment any of our existing structures to reflect this? Both these questions appear to us to merit furtner thought during the course of selecting ESD indicators.

3 Indicators as Objectives, Inputs, and Constraints

3.1 Introduction

To this stage, we have been fairly abstract in our treatment of indicators within the decision process. Although a number of different decision processes have been described, we have not yet described how the role of any specific indicator can differ. To that end, we shall now be more specific about the role of indicators, although we shall leave behind the distinctions we have made between different decision processes.

Our primary purpose here is to provide a framework adequate to describe a generic decision process, which will allow us to highlight some of the key roles that indicators play.

In our descriptions of the collective decision process, we have often alluded to the fact that we will also have to compare these to the preferences of individual agents. In this section, we shall for generality regard an agent as any single entity, such as an individual consumer, a cooperative group, an individual firm, an industry, a government institution, or indeed, society as a whole, that can take information and use it within some decision process.

Before proceeding with a description of the generic decision process, however, it is useful to bridge the discussion of collective decisions and individual decisions with an example of how the individual decision process might affect information requirements in the collective decision process. It was shown in some cases, that a social planner required information relating not only to environmental goods and services, but also to individual and community preferences for those goods and services.

Recent literature has provided an active forum for addressing the appropriateness of different preference measures. The discussion has concentrated on two measures – willingness to pay (WTP) and willingness to accept (WTA). The WTA measure provides an indication of how much income compensation individuals would accept for the loss of some good or service, whereas the WTP measure estimates how much income they would be willing to give up in exchange for some benefit. Welfare economics shows that WTA is the appropriate measure when valuing losses in environmental services or losses in amenity. Although in theory the difference between WTA and WTP is normally small when evaluating price changes (Willig 1976), empirical evidence (Knetsch and Sinden 1984; 1987) has shown that WTA can exceed WTP by up to an order of magnitude. This implies, for example, that an individual might be willing to pay $10 to have an extra tree to look at on the boulevard, yet might require $100 in compensation if one of the trees presently on

Institutional Structure or System								
	No Gov't	Systems with Property Rights and Contract Enforcement						
		Minimal Gov't	Reform Government		Distributive Government		Reform and Distributive Government	
		Unanimity	Dictator	Democracy	Dictator	Democracy	Dictator	Democracy
Relative Performance of System: * Poor ** Satisfactory *** Very Good								
Efficiency of System	*	* * *	* * *	* * *	* *	* *	* *	* *
Equity of System	*	* *	* *	* *	*	* * *	*	* * *
Liberty (Intrusions)	*	* * *	*	* * *	*	* * *	*	* * *
Liberty (Property Loss)	*	* * *	*	* *	*	* *	*	* *
Relative Indicator Requirements of System: * Low ** Moderate *** High								
	None	*	* * *	*	* * *	* *	* * *	* *

Figure 5: Summary of Effectiveness and Requirements of Collective Decision Structures

the boulevard were to be removed. It might be argued that this discrepancy is simply a measurement error related to the particular measurement techniques, but there is some theoretical basis for this discrepancy.

Theoretical economics literature provides one such rationale. As Hanemann (1984) illustrated, the difference between WTA and WTP could be large if there are no or few private substitutes for environmental goods or services. In addition, the result is not anomalous to psychologists who observe that people are usually more adamant about keeping what they have than about getting what they do not have. Often, this is associated with the belief that they have innate usufructuary rights to – or implicit ownership of – some existing, albeit publicly available, asset.

If, as evidence suggests, WTA does indeed exceed WTP for environmental goods and services, then such an asymmetry can have important implications for policy implementation. A case in point would be the tradeable emission rights discussed previously. The actual final emission levels would no longer be independent of how the permits were distributed; one might well imagine that considerably less pollution would occur if the rights were given originally to the local Fish and Wildlife Association rather than to the industry, even in the presence of trading.

From a practical point of view, attempts to value untraded environmental goods usually rely on techniques that measure WTP. Indicators based on such techniques may understate individuals' preferences and some adjustments would be necessary. Because of the empirical difficulties with measuring WTA, such adjustments may have to be made subjectively.

3.2 The Generic Decision

Irrespective of who or what is making a decision, economists generally characterize a given agent's problem as consisting of three primary components – objectives, inputs, and constraints, as shown in Figure 6. In general, the problem consists of some objective that the agent attempts to optimize. The agent reacts to indicators and, in so doing, adjusts actions to achieve that objective. The indicators will normally relate to either the inputs to or constraints on the decision process. In some cases, indicators may also measure the actual objective. In individuals, this objective is often supposed to be some form of utility that is not directly measurable, whereas for a firm it would be profitability, which is measurable. In addition, there will also be some set of available information that is not relevant to the decision. We include it for completeness primarily because, as noted earlier, some of the less efficient collective decision processes also require individual preference information about irrelevant alternatives.

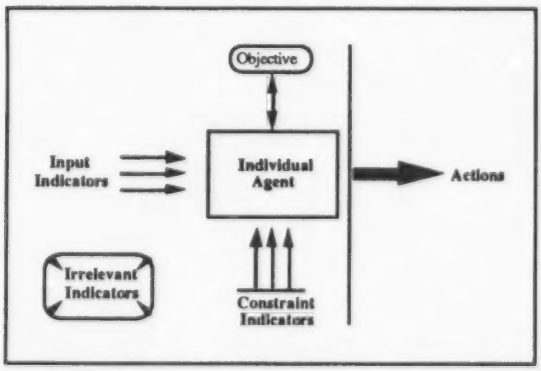


Figure 6: Generic Decision

The reader will note that we have not explicitly attached any indicator to the action of a given individual agent. That is because such an indicator would only become important if it actually formed part of the input information, constraint, or objective of some other decision agent, or it becomes part of the set of irrelevant information.

Although this is a relatively simple framework, it is very versatile and can be used to incorporate any of a host of inputs or constraints. It is premised on what is often referred to as a utilitarian decision process and is mathematically tractable for such processes, but it can be used conceptually for most political decision processes. It readily handles traditional utility or profit-maximizing agents; it can handle the Rawlsian system of justice (Rawls 1971) which involves maximizing the well-being of the least fortunate person; and it can handle, at least conceptually, even those systems based almost purely on libertarian or ethical grounds, such as those described by Nozick (1974) and Buchanan (1975).

On the surface, it may appear that this framework does not handle multiple objectives. This relates to the idea that, in many decision problems, we are trying to satisfy more than one goal – economic equity, economic growth, environmental integrity, civil liberty, and so on. But in this framework, such problems require us to think of any conflicting goals in terms of potential trade-offs (Griffin 1977) that might be necessary to achieve some greater ultimate objective, or where no such trade-offs are possible, we are forced to regard as constraints at least some of what we thought were goals. In this manner, some single objective is always dominant.[5] For example, sustainable development may be a dominant objective for which economic growth or ecological integrity are regarded as inputs.

3.3 Indicators as Objectives

An indicator as an objective is meant to reflect how well we are achieving a given target. If, for example, we are interested in species diversity, we may wish to measure some species diversity indicator and keep track of it as we implement policies intended to improve diversity. A relevant issue in the design and use of indicators in such a role relates to whether we are trying to optimize or simply to improve. In most cases, we tend to assume that the two amount to the same thing, but this is not generally true.

Consider, for example, the case shown in Figure 7. The objective indicator (ecological integrity) is on the vertical axis and some policy

[5] Failure to find a dominant objective means that the decision process fails on Arrow's axiom of universal domain (U) or rationality (R) to the extent that possible combinations exist over which no preference can be stated.

variable is on the horizontal axis. Suppose our goal is to maximize ecological integrity; this requires that we implement policies at some level P^* and achieve the maximum ecological integrity I^*. If we happen to start at point P_1 and seek small improvements in the objective indicator, we will eventually achieve this maximum. If we happen to start at point P_2, however, and seek small improvements in the objective indicator, we will eventually end up at the local maximum I'. We will not have achieved maximum ecological integrity I^* and we will have headed off into the wrong policy direction; the correct policy direction would have been to accept some short-term degradation in integrity in the interest of longer-term improvements.

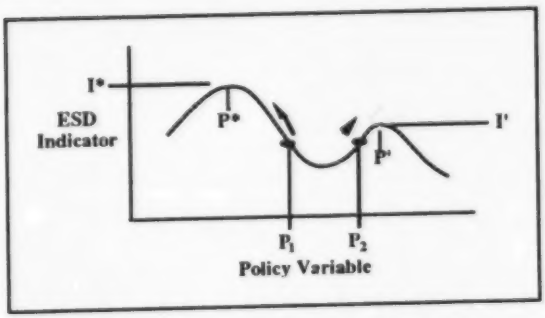


Figure 7: Optimizing Ecological Integrity. Where are we and which way is up?

The implication this has for reporting indicators is that if we pay strict attention to the gradients or changes in indicators, we may be making the wrong policy choices if our objective is to maximize that indicator. If we are seeking to maximize some indicator, we will more generally require forecasts of that indicator under a range of policy alternatives rather than simply a forecast of how an indicator will change in response to a small policy perturbation.[6]

[6] This same phenomenon is responsible for limiting the application of cost-benefit analysis to small projects: those that do not themselves affect shadow prices. Shadow prices, in fact, measure the slope of the objective function, and CBA manages to find projects or policy reforms that push the economy up that hill. Although in theory it is possible with CBA to identify the highest point (see Hammond 1980; 1983), this requires more extensive global knowledge than that available simply from the gradients of the objective function.

3.4 Indicators as Inputs to the Objective Function

An indicator as an input is meant to act as one factor that we are trying to consider in achieving some objective. We are willing to accept trade-offs of one input with other inputs. This suggests that one of the key roles of analysis is identifying what kind of trade-offs exist between various inputs. In economic terms, economists usually concentrate on whether inputs are complementary or substitutes.

In terms of indicators as inputs, the corollary we must address is whether two given indicators will generally move in the same or opposite directions. The debate between economic growth and environmental integrity is becoming a classical example of this. Some argue that they are complements to the extent that a strong economy is needed to guarantee environmental integrity, and environmental integrity will in turn lead to a healthy economy. Others argue that these are substitutes; one can be achieved only by sacrificing the other. Pearce and Turner (1990) hypothesized that the degree of complementarity is actually income dependent; growth and integrity are complements in low-income countries but substitutes in high-income countries. The debate will no doubt continue, but it underlines the importance that where indicators are used as inputs to the decision process, we must try to formulate them in a way that will allow us to determine what the trade-off or synergistic opportunities are between inputs.

3.5 Indicators as Constraints on the Objective Function

Constraints on a decision system are in a sense among the most interesting, in that they can originate from many sources. They can reflect natural limits, such as the minimum food intake a person must have to survive. They can reflect institutional limits, such as the regulations and laws that govern us. Or they can reflect moral or ethical considerations that constrain our activities to those that are ethically acceptable; Partridge (1981) and Norton (1988) argue that certain rules of ethics should be treated as constraints to defining the feasible set of alternatives within any given decision or allocation process. To Stone (1972), for example, trees and other natural objects should have legal rights such that we would no sooner eliminate a species of tree than we would murder an individual.

What distinguishes constraints from other inputs is the lack of trade-off opportunities beyond

some critical value. As noted above, these constraints can be self-imposed or physical, but in any event they are not allowed to pass the critical value.

A relevant example of the different roles of ESD indicators as constraints or objectives again relates to the roles of conventional economic growth indicators and ecological integrity indicators. If we have determined, for instance, that economic growth and ecological integrity are, in fact, substitutes, two of many alternatives we may have for setting policies and making decisions are:

a) maximize economic growth subject to the constraint that ecological integrity remains constant or does not fall below a certain level; and

b) maximize ecological integrity subject to the constraint that economic growth remains constant or does not fall below a certain level.

3.6 An Example

To illustrate how a single indicator can be treated and used differently in a decision process, we shall consider an example that relies on a visual model often used in renewable resource economics. We shall first describe the model and then show how it leads to four different results for different roles played by the information in the model.

We shall consider a typical harvesting problem for a renewable resource and ask what the equilibrium sustainable rate of harvesting is. A useful technical model for approaching this problem is shown in Figure 8. It is commonly used in modelling biological populations, especially in fisheries (Clark 1976).

The simple form of these types of models – Figure 8(a) – shows population growth rate dX/dt as a function of population stock size X. The reader is reminded that time t is not explicitly shown on either axis; growth rates start out low when population size is small and gradually increase to some maximum before falling off again.

Figure 8(a) thus illustrates a case where a population will grow to a maximum stock size of $\bar{x}$. A plot of population size versus time would show that $\bar{x}$ was never actually reached, but that it was approached asymptotically.

There are two characteristics of interest in this model. First, the population size is self-limiting because of some external constraints on the system, often food or habitat availability. In the biological literature, this characteristic is known as compensation. Second, a process of depensation occurs at low stocking rates. This reflects the fact that populations often exhibit low initial growth because of higher vulnerability to predation or lower breeding opportunities for sparsely distributed species; here, it generates an inflection in the growth function. In the model, a population can be maintained at any given size X in perpetuity by choosing a harvesting rate equal to the growth rate at that stock size. The maximum sustainable yield (MSY) is said to occur at a population size (X_{MSY}) at which dX/dt is at a maximum. Depending on the nature of the harvesting cost function, optimal sustainable economic equilibria that maximize profits alone can be generated anywhere on this function.

Although the model in Figure 8(a) is often used, Figure 8(b) is often asserted to be a more realistic depiction of certain types of populations. It is characterized by a process of critical depensation once the population falls below a critical size X_C, it will eventually become extinct because its natural growth path experiences an annual decline. For specific wildlife resources this may arise from minimum stocks necessary for viable breeding, or from increased vulnerability to natural fluctuations in the environment.[7]

In entire ecosystems, the model in Figure 8(b) would appear to capture three of the salient features often encountered in the technical literature. First, it reflects that mature climax ecosystems appear to be extremely stable in their virgin state, characterized by the equilibrium population $\bar{x}$. Second, it captures the idea that some form of harvesting of the ecosystem would be sustainable between x_c and $\bar{x}$. Finally, in the area of critical depensation, it shows that

[7] Strictly, one might fairly ask how a population evolved beyond X_C in the first place if it presumably started at a point close to zero. The usual simple response asserts that over long periods of time the growth function itself changes as a system evolves within its environment. One might imagine, for example, that a growth function of the sort in Figure 8(a) applies at an early stage in the process such as before an equilibrium population is reached, and that 8(b) applies after the equilibrium is achieved. Such an explanation is also consistent with observations that future changes can depend on past conditions; a growing population will not necessarily follow the same growth path as a declining population.

if the ecosystem is over-harvested it will be incapable of regeneration and will continue to decline in quality even if harvesting is halted. The point at X_C is mathematically a bifurcation and, in some cases, equivalent to what Kay (1990) would call a catastrophe threshold. In the absence of any stresses, $\bar{x}$ will be a system's 'optimum operating point'.

The technical problems associated with describing such a growth function are beyond the scope of this paper. We now turn our attention to how this information could be used in different decision contexts.

Indicators as Objectives: A simple example involves trying to maximize the actual sustainable physical yield from a given resource. This is illustrated at point MSY in Figure 8, corresponding to the peak of the growth function. The equilibrium population level in this case is X_{MSY}.

Achieving this point requires in theory that we know what the entire growth function looks like if we are looking for the global optimum point, or if we are content with finding a local maximum as discussed earlier, we require information only in the vicinity of our current position on the growth function.

Indicators as Inputs: The usual example here is to maximize either social or private profitability from some resource given economic inputs, such as output prices, tax structures, interest rates, and harvesting costs, and technical characteristics, such as the curve shown in Figure 8. In this case, the technical information acts as an input to the harvesting decision, which will typically, but not necessarily, result in equilibrium harvesting levels somewhere between population levels X_C and X_{MSY}.

Again, achieving this point requires in theory that we know what the entire growth function looks like if we are looking for the global optimum point. In practice, however, we are generally concerned with information only in the vicinity of our current position on the growth function. We might be concerned with getting more information if it is determined that the final operating decision is very sensitive to this input (we shall discuss this point later in Section 4). In addition, it is clear that just knowing the growth rate (dX/dt) will not necessarily tell us where on the growth function we are; we also need to know population size (X). This might seem like a trivial point, but failure to obtain such key information was thought by Paulik (1971) to contribute to the collapse of the Peruvian anchovy fishery – managers assumed they were at a population size to the right of MSY when in fact they were to the left of it.

Indicators as Ecological Constraints: Kay (1990) has described catastrophe thresholds as points at which ecosystem change becomes irreversible. Although Kay does not take the position that such thresholds should be regarded as constraints, Bishop (1978) asserts that typical optimization procedures are not helpful in practice and that where the possibility of extinction arises, development should be at levels that set a safe minimum standard for populations and that no amount of development

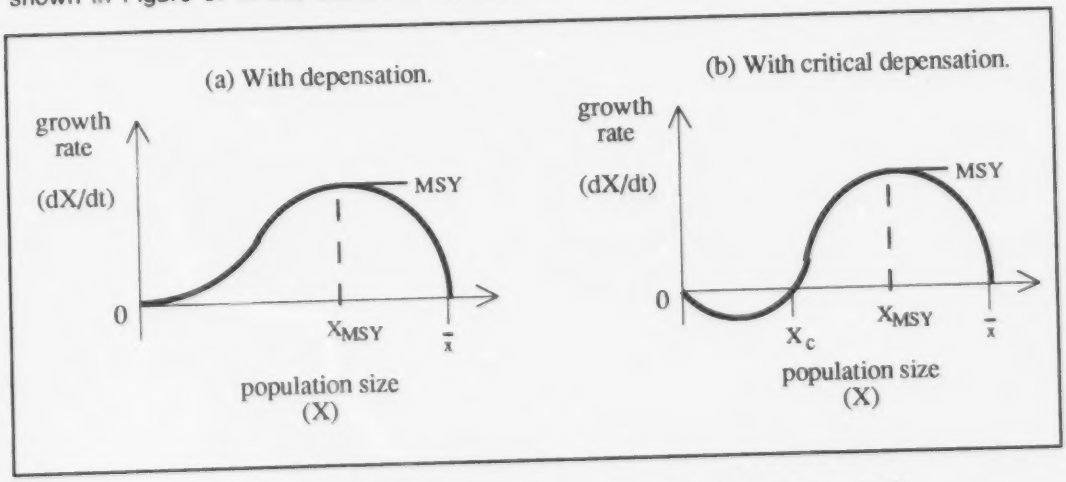


Figure 8: Population Models. To reap or not to reap? (and if so, how much?)

should be pursued if these standards might be breached. In essence, the catastrophe threshold becomes a constraint; development would be allowed to a level of the safe minimum standard at a population level X_C. But we have seen that other procedures would under some circumstances generate results that allowed population levels to exceed X_C. If we treat indicators as constraints, we would be more concerned with knowing how the indicator behaved at the critical threshold and less concerned with how it behaved at the optimum.

Indicators as Moral Constraints: As a final example, we consider the extreme moral case (Stone 1972, for example) which requires that animals and plants have rights and that as a moral prerogative, we should not kill anything. In such cases, we would get the result that the resource is not harvested and will eventually be left to establish an equilibrium at $\bar{x}$. In cases where such constraints apply, we require primarily indicators that would effectively allow us to monitor harvesting activity, or failing that, the size of the resource stock.

3.7 Summary: Implications for Indicators

This section reviewed the various roles that indicators could take in the decision process. A generic decision process was described that requires information on indicators as objectives, inputs, and constraints. Within the context of ESD indicators, some notable conclusions arise:

a) If ESD indicators are used as objectives, it is not necessarily valid to look at only directional changes in these indicators. A small improvement in the indicator does not necessarily mean we are heading in the right direction.

b) Using indicators as different factors in the decision process will affect both the outcome of the decision process and the degree of detail required of the indicator. Indicators as objectives tend to require the most detail, indicators as constraints tend to require the least detail.

c) Given the substantial effort required in collecting indicator information, it is suggested that most of the efforts in this area concentrate on measuring indicators that will be useful as constraints to the decision process.

4 Indicators, Expectations, and Risk

4.1 Introduction

The foregoing discussion has assumed that decisions can be made under some degree of certainty. Most policy analysts recognize, however, that the information we do have available is subject to considerable uncertainty. We may be able to measure low-level sulphur emissions, for example, but there is still little evidence of what effect, if any, such emissions have on the very long-term health of humans. Similarly, while we can often foresee some of the immediate environmental consequences of a single project, the cumulative longer-term consequences of several activities are often not clear. Such problems are endemic to the decision processes of both private individuals and government planners.

In this section, we shall highlight some of the major conclusions of the risk literature, concentrating on those most relevant to sustainable development.

Although people tend to refer colloquially to risk and uncertainty interchangeably, the formal approaches to this subject usually distinguish clearly between the two. Whereas risk is something that is generally regarded as quantifiable, through the assignation of probabilities, for example, uncertainty is not quantifiable. Approaches dealing with the two are thus quite different. Uncertainty is generally handled using analytical techniques that do not rely on any quantification; risk is generally handled using some form of mathematical modelling that assigns numbers to various inputs, constraints, or objectives in the decision process.

Before proceeding with risk proper, we shall first address some related issues of how we might make decisions under conditions of uncertainty.

Our discussion of risk will then concentrate on some analytical methods of dealing with risk. These can, in fact, be divided into two sets of economic modelling approaches – those dealing with the expected case and those dealing with the actual risks of deviating from this expected case. Our discussion draws on a review of 18 methods of dealing with the expected case. We shall then briefly review four methods of risk analysis typically used within economic analyses, with a view to determining what, if any,

supplementary information may be required of indicators.

4.2 Addressing Uncertainty: Lessons from the Literature on Option Value

Irreversibility consistently stands out in the literature as an important issue that tries to address uncertainty. While uncertainty relates to the fact that we simply do not know what the consequences might be of any given action, irreversibility relates to the notion that whatever those consequences are we will not be able to return to the original state. As discussed by Kay (1990), for example, ecosystems undergoing catastrophic changes will not return to their original state. Any options afforded by the original state are forever lost. Since one of the founding definitions (WCED 1987) of sustainable development included the idea that we should keep future options open, it is useful to review the economic basis of option value.

The existence of a positive option value related to certain environmental resources is often regarded as a theoretically and empirically important valuation issue that is usually overlooked in the process of resource valuation. Option value essentially intends to capture the amount that an individual would be willing to pay to preserve the option to have access to an environmental good or service.[8] The essential observation is that individuals would pay to keep this option open, even if they did not have any specific intent of visiting the area that might provide the good or service. In addition, society as a whole may be willing to pay to have access to a specific genetic resource at some time in the future, even if it has no immediate consumption or production value. This value, that of being willing to pay for the option of having certain access to consumption, we describe as basic option value. It is normally implied that this consumption will be at some future date, although Smith (1983) notes that the passage of time is not essential to the formulation of this type of option value.

Gallagher and Smith (1985) subsequently defined access option value as the amount that individuals would be willing to pay for increasing the probability that they would at one time have access to a given environmental good or service. In essence, they simply recognized that even once an option was purchased it might not guarantee the ability to exercise that option. Nonetheless, it seems clear that individuals might be willing to pay for an improvement in their opportunity set, which is essentially what access option value reflects. As with the basic option value, the existence of access option value does not depend on the passage of time.

Freeman (1984), by noting that irreversible decisions were often involved in environmental projects, argued that a quasi-option value existed in conservation projects. This argument essentially extended the time-dependent aspects of option value put forward independently by Arrow and Fisher (1974) and Henry (1974b). It requires that, at the outset, the planner does not have complete information on the intrinsic value of the resource. In the case of a wetland area, for example, the argument would be that we do not know everything about the wetland's value. If we destroy it through development, we will never find out. If we defer development, then the opportunity that we have generated for finding out more about the intrinsic value of the wetland is our quasi-option value. Fisher and Hanemann (1985) proved that this quasi-option value is positive, provided the expected value of the information that might be gained by deferring development is positive.[9]

In general, therefore, the option value arguments assert that valuing *in absentia* benefits and incorporating them within a decision process are potentially important tasks.

4.3 Addressing Uncertainty: Lessons from the Literature on Decision-Making

Environmental and welfare economics literature have for some time been preoccupied with how decision-makers should make decisions under conditions of uncertainty.

The earlier literature dealing with uncertainty formalized its notation by regarding each possible state of nature to be a different attribute of a given commodity (Debreu 1959). This allowed uncertainty to be treated in a typical utilitarian framework such as that described in

---

[8] The earlier papers on this topic include: Weisbrod (1964), Cicchetti, Seneca, and Davidson (1969), Cicchetti and Freeman (1971), and Fisher, Krutilla, and Cicchetti (1972a; 1972b; 1974).

[9] Hanemann (1989) has more recently argued that, under certain circumstances where a continuum of development levels is available, the quasi-option value might, in fact, be negative for some levels of development. This is analytically somewhat worrisome, and it seems more reasonable to interpret quasi-option value as zero in such instances.

Section 3. The probabilities of various states of nature and conditions under each state simply became additional inputs to the decision-maker's optimization problem.

Subsequent literature investigated the implications of uncertainty on optimal investment decisions, and in two notable cases it was shown that the treatment of uncertainty by the collective was rightfully different from that of individuals. Samuelson (1964) argued that, through what can be described as risk-pooling – where individuals can pool risks by combining many risky projects – the collective will be apparently less risk-averse than any one individual and, will, in aggregate, invest more in the current period.[10] The same conclusion – that society should invest more than individuals – was made in a seminal article by Arrow and Lind (1970), which argued that uncertainty could be ignored under certain conditions. They proved that, in evaluating public projects, individual risk premiums (reflected in private discount rates) vanished in the limit, and random variables could be replaced by their expected values if private and public risks were uncorrelated and if the public risks were absorbed by a large number of individual agents. This simplified decision-making tremendously, since rather than having to consider all of the probabilities of the various states of nature, one need consider only the expected state of nature.

This risk-spreading assertion by Arrow-Lind resulted in a substantial amount of theoretical debate on the nature and effects of uncertainty in economics literature in general[11] and environmental economics literature in particular. In much of the discussion of option values, and especially in articles by Schmalensee (1972), Fisher (1973) and Smith (1987), the Arrow-Lind hypothesis is attacked and shown to be of limited validity when an option demand exists for a good or service. It is further argued by Arrow and Fisher (1974) and, more recently, by Viscusi (1988), that when investment is irreversible or when it has irreversible environmental effects, optimal investment decisions would involve less investment than might otherwise occur in the absence of such irreversibility.

In short, although the generalizations relating to uncertainty provide useful insights, they provide little guidance as to the actual levels of investment that are optimal other than, as noted by Fisher (1989; p. 86), "to go slow."

4.4 Some Practical Methods of Risk Analysis: The Expected Case

The first stage of any practical risk analysis – and at times the only stage – inevitably involves describing what we call the expected case. This expected case is usually meant to represent our best estimate of some input or constraint in our decision process.

In this context, the role of analytical modelling is becoming increasingly important as a decision-making tool. Figure 9 lists 18 approaches to economic analysis sometimes used to assist the decision process. Appendix B summarizes these approaches, each of which effectively provides the decision-maker with information relating to the expected case.

An exhaustive and comprehensive review of these approaches is not possible here. They have, however, been separated into five categories to facilitate discussion and to highlight some salient features relevant to ESD indicator design and selection.

Preference Description Approaches
Contingent Value Methods (CVM)
Hedonic Price Methods (HPM)
Limited Dependent Variable Methods (LIMDEP)

Structural Description Approaches
Technical Input-Output (IO-T)
Economic Input-Output/Fixed Coefficient (IO-EF)
Economic Input-Output/Variable Coefficient (IO-EV)

Impact Description Approaches – Static
Input-Output (IOIO-EF & IO-EV)
Partial Equilibrium (PE)
General Equilibrium (GE)
Mathematical Programming (MP)

Impact Description Approaches – Dynamic
Optimal Control Approaches (OC)
Statistical Time-Series Analysis
Econometric Analysis
Structural Methods

Impact Evaluation Approaches
Cost-Benefit Analysis (CBA)
Cost-Effectiveness Analysis (CEA)
Policy Reform Analysis (PRA)
Policy-Effectiveness Analysis (PEA)

Figure 9: Economic Analysis Mechanisms

[10] In the literature dealing with discount rates, this is also taken as an argument for using social discount rates, which are lower than private discount rates.

[11] See, for example, Henry (1974a, 1974b), Schmalensee (1976) and Sandmo (1972).

Preference description approaches attempt to provide indicators of what individual preferences are. Where competitive markets exist, these preferences are usually presumed to be reflected by prices. In the absence of markets – as with most environmental goods and services – these techniques are intended to impute some value of individual preference to untraded commodities.

Many of these techniques have been very well developed. As they are intended to address primarily untraded goods, much of the theoretical and empirical work that has been done involves environmental issues. Appendix B notes that generally such techniques can have quite different information requirements. One of the most promising little-used approaches involves techniques that use qualitative variable (LIMDEP) approaches; these tend to be very information-efficient while still providing useful preference information.

Structural description approaches describe the linkages between various sectors of an economy in either physical terms (IO-T approaches) or monetary terms (IO-EF and IO-EV). These types of approaches are often loosely referred to as input-output models, distinct from other types, but they often represent a building block for other approaches.

There are really two major points about these approaches that are relevant to ESD indicators. First, they underline the distinction between technical flows and economic flows. In simple terms, we are reminded of some general identity of the type:

Value = Price x Quantity

The economic approaches tend to concentrate on value, whereas the technical approaches focus on flows or stocks of physical quantity. If environmental goods are to be considered within an economic framework, they will often require that some value be imputed to them. While techniques exist for imputing such values,[12] the relevant approach will depend to a great extent on how one wishes to regard environmental goods theoretically. Victor (1990) outlines the difficulty in coming to agreement, for example, on how capital in general and environmental capital in particular should be treated.

Second, the review illustrates that technical models of environmental quantities are often modelled in isolation of economic structural models. It is generally evident in the literature that significant potential exists for integrating some of these approaches.

Impact description approaches are distinct from the above approaches in that they take the analysis one step further – they are at least partly predictive rather than entirely descriptive. The static approaches generally attempt to describe what will happen within a single or indeterminate time period, whereas dynamic approaches are essentially forecasting tools.

The simplest static approaches actually depend on shocking the IO models to predict, using the existing structures, what the impacts will be on some physical or economic indicator. More complex approaches impose some equilibrium on the system. Equilibrium is achieved through allowing prices to vary either in one sector (in PE models) or in the entire structure of the economy (in GE models). Techniques with the greatest potential appear to be those that integrate some of the technical and economic input-output approaches. PE models also hold potential for evaluating selected environmental goods and services.

The review of dynamic forecasting methods suggests that most of these approaches will not be useful until we develop the indicators required for static impact description. The one exception to this is time-series analysis, but this technique is generally not useful in environmental applications because it requires long historical time series, and even then is primarily useful for short-term forecasting.

Impact evaluation approaches take the impact analysis through to the final step – they use the impacts predicted from the above approaches to evaluate the desirability of a given project (in the case of CBA and CEA) or policy (in the case of PRA and PEA). These approaches have an explicit objective, the improvement of social welfare, which requires that all current and future physical and financial flows are reduced to a single indicator. This indicator is then used to determine whether the given project or policy improves social welfare.

These approaches have been used with moderate success in analysing projects with environmental impacts, but the policy evaluation approaches have generally received little

---

[12] Johansson (1987) provides an excellent review of many of these specific approaches.

attention. As PRA and PEA are, in principle, simply extensions of the same techniques used in CBA and CEA, it seems that they should provide a natural complement to analysing impacts.

A major constraint to the four impact evaluation approaches is that they identify changes that are beneficial but, as noted in Section 3, they do not necessarily move us toward the optimal allocation of resources.

4.5 Some Practical Methods of Risk Analysis: Departures from the Expected Case

The second stage of any practical risk analysis involves describing what would occur if the expected case did not, in fact, occur. Appropriately, we ask whether any analytical techniques might be applied to assist decision-makers in selecting policies in the face of risk. It is interesting to note that most decision-makers use no formal tools for risk analysis but rely on judgemental approaches. Ruitenbeek and King (1988) note that even when such tools are available to them, planners tend to follow their judgement if they obtain contradictory signals. Nonetheless, some formal methods are available.

Within the context of resource and project evaluation problems, we can characterize four risk analysis approaches that are at times used to quantify risk. In order of increasing information requirements and complexity, they are summarized in Figure 10. Appendix C reviews each of these approaches in further detail.

To a large degree, these approaches are meant to illustrate the types of methods that can be used as add-ons to many of the impact, forecasting, or evaluation approaches discussed in the previous section. They are not, as such, normally constrained to only selected types of modelling.

Discount Rate (DR) Approaches
Ad Hoc Sensitivity (AHS) Analysis
Critical Value Sensitivity (CVS) Analysis
Probability-Based Sensitivity (PBS) Analysis

Figure 10: Risk Analysis Approaches

Discount rate (DR) approaches generally involve adjusting discount rates in project evaluation processes in a direction that favours maintaining environmental integrity. Project analyses (such as CBA and CEA) require that an explicit identification of preference between current and future benefits be expressed. The discount rate is the economist's short-hand method for comparing such benefits. A discounting process typically values future benefits and costs less than current benefits and costs. Selection of an appropriate discount rate will normally reflect the opportunity cost of real investments in an economy, as well as the rate of time preference displayed by governments and individuals through their savings functions. It is often argued that social discount rates should be less than private discount rates because of intergenerational equity issues (Marglin 1963) or risk-pooling issues (Samuelson 1964), and that environmental investment projects should have lower discount rates attached to give explicit additional weights to future benefits.

The major advantage of this approach is that it requires essentially no additional information in undertaking evaluations, but the use of such a strategy is usually questioned because, as noted by Markandya and Pearce (1988), among others, there is no particular reason for believing that high discount rates or low discount rates will be more environmentally friendly. High rates tend to discourage or quash conventional investment, which might have been detrimental to the environment. But low rates can promote projects with substantial future environmental benefits. There is no *a priori* reason for predicting which way the bias lies, unless all the environmental details on the project are available. In that event, however, the analysis would be superfluous.

Ad hoc sensitivity (AHS) analyses are commonly used to determine how robust a decision is to slight variations in the input information. The basic question we are asking ourselves is: Supposing that some indicator is actually different from what we expect it is, would that difference be enough to affect our decision? Suppose that ecological studies predict that a given project is likely to eliminate 20% of fish stocks because of habitat losses, and that we regard this loss as acceptable and decide to go ahead with the project. An ad hoc sensitivity test would ask this question: Suppose the losses were 25%, would we still go ahead? The information requirements for such a process are fairly minimal; we require just an estimate of the expected impact of the project. The

disadvantage is that there is no analytical basis for specifying the sensitivity ranges.

Critical value sensitivity (CVS) analyses are less arbitrary than the AHS methods in that they relate the decision to some objectively determined critical value. The most common and simplest of CVS involves simply calculating those input values at which the decision actually changes. Extending the earlier example, we ask how much fish loss we are willing to tolerate before we decide not to go ahead with the project. If we determine that we will tolerate up to 50% fish loss, for instance, we are then in a position to go back and determine how likely such an event will actually be.

The advantage of this approach is that it provides some indication of how close we are to practical ecosystem limits. The disadvantage is that it places essentially no weight on good outcomes. In that sense, it is similar to the 'maximin' strategies, such as those of Rawls, that concentrate on selecting the strategy with the least detrimental worst case.[13] In the case of environmental commodities, however, it may be a relevant approach.

One example of a CVS test involves calculating an index of gambler's ruin. The gambler asks himself how many times he can afford to roll the dice before he goes broke, assuming that he loses every time. Projects with low indices are risky; those with high indices are safe. What is interesting about this example is that, although it is traditionally acknowledged that individual agents might make some decisions involving such an index, it is not normally regarded as a proper approach for society as a whole to take. Because of risk-spreading, the concept of going broke was not a valid one for social choices. In the context of doing irreversible damage to critical environmental resources, however, we may well wish to ask how many times we can afford to roll the dice.

Probability-based sensitivity (PBS) analyses are the most information intensive, in that they require the analyst to specify precisely what the chance will be of any given outcome. In our fisheries example, we must ask what the probability is of 0% fish losses, 10% fish losses, 20% fish losses, and so on up to 100% fish losses.[14] We then use this information to determine what our action will be.

In practice, such approaches are hampered by a quite significant technical requirement of providing all of this information. A related problem is that the probabilities assigned to any given event are often dependent on subjective judgement and are not reliable in the context of a plurality of views.[15] Empirical testing of inputs into Monte Carlo procedures, for example, show that people tend to assign too much weight to the most likely values and underestimate the amount of variation that exists. Furthermore, where surprises are common, as is often the case in ecological systems, there is little empirical basis for assigning probabilities; we quickly pass from a situation of quantifiable risk to one of unquantifiable uncertainty. The approach has theoretical appeal but severe practical limits.

In short, it is the author's opinion that the most tractable solution to dealing with risk in the context of ESD indicators is to rely on the CVS approach.

4.6 Summary: Implications for Indicators

This section addressed some of the theoretical and practical aspects of incorporating risk into the decision process and determining what requirements this may have on the selection of ESD indicators.

The conclusions arising from this discussion that are directly relevant to indicator design are:

a) Because of the existence of option value, which is often related to the irreversibility of decisions and their consequences, we may wish to produce indicators reflecting this option value.

b) In a brief review of analytical approaches useful for modelling the expected values of inputs or constraints to the decision process, formal models that

---

[13] There is substantial literature in economics that relates to games theory. The approaches used in games theory essentially attempt to describe individual decisions as part of a gaming strategy, which is intended to maximize some given objective. The maximin strategy is one of many often discussed. An excellent volume relating to economic analysis in this context is that edited by Feltham, Amershi and Ziemba (1988).

[14] In practice, this is usually simplified somewhat by reporting an expected value m and standard deviation s, with a coefficient of variation – defined as the ratio CV=s/m – providing an index of risk. Low CV has low risk.

[15] Clear examples are provided in Victor (1990) and Kay (1990). Victor illustrates the different economic opinions of how capital should be treated, and Kay discusses the different ecological methods of measuring ecological integrity.

integrate both ecological and economic considerations have the greatest potential. At a minimum, this requires information that describes some of the key linkages between the economy and the environment. As a next step, it requires information that can be used as ecological constraints in the economic models.

c) A requirement in designing many indicators will be to link environmental and economic systems. In practice this will normally require some monetization of the flows of environmental goods and services, as well as the stocks from which they are derived.

d) In a brief review of risk analysis approaches, those relying on the use of extreme values or critical switching values are probably the most tractable and relevant to the issues of sustainability. In practice, this will require reporting structures of indicators that compare expected or existing values with some critical value that is regarded, for example, as a catastrophe point for an ecosystem.

5 Concluding Remarks

A recurring theme throughout this paper has been that the decision process and the techniques it uses can have a bearing on the type of ESD indicators that are eventually required. As specific summaries relevant to indicator selection have been provided at the end of each section, we shall close with some remarks that are not directly related to indicators, but which relate to the actual analytical structures that are likely to use and produce such indicators – economic and ecological modelling.

It is clear that economic models are increasingly becoming a component of the decision process. They are used to describe existing structures, predict impacts based on those structures, evaluate the desirability of those impacts and assess the importance of various types of risk and uncertainty.

Given the propensity to use such economic models, a new role is emerging for ecosystems modelling as an important complement to economic modelling. Although there has been some limited experience in integrating economic and environmental factors into single models, it seems inevitable that if applied economists are to contribute meaningfully to discussions of ecologically sustainable development, such integration will continue. At the least, economists will familiarize themselves with some of the methods and conclusions arising from ecological modelling. The next stage will then involve testing ways of integrating some ecological and economic approaches to arrive at models that, although imperfect, will provide a basis for further improvement to the art of both ecological modelling and economic modelling.

References

Ahmad, E., and N.H. Stern. 1987. Alternative Sources of Government Revenue: Examples from India, 1979–80. In *The Theory of Taxation for Developing Countries*, by David Newbery and Nicholas Stern (eds.). Washington: Oxford University Press.

Arrow, K. 1963. *Social Choice and Individual Values*. New York: J. Wiley & Sons.

Arrow, K., and A.C. Fisher. 1974. Environmental Preservation, Uncertainty, and Irreversibility. *Quarterly Journal of Economics* 88:312-319.

Arrow, K., and R.C. Lind. 1970. Uncertainty and the Evaluation of Public Investment Decisions. *American Economic Review* 60:364-378.

Bishop, R.C. 1978. Endangered Species and Uncertainty: The Economics of a Safe Minimum Standard. *American Journal of Agricultural Economics* 60:10-18.

Block, W. (ed.) 1990. *Economics and the Environment: A Reconciliation*. Vancouver: The Fraser Institute.

Bradley, P. 1979. Production of Depleting Resources: A Cost Curve Approach. Vancouver: Department of Economics, University of British Columbia.

Buchanan, J. 1975. *The Limits of Liberty*. Chicago: Chicago University Press.

Cicchetti, C.J., and M. Freeman. 1971. Option Demand and Consumer Surplus. *Quarterly Journal of Economics* 85:528-539.

Cicchetti, C.J., J.H. Seneca, and P. Davidson. 1969. *The Demand and Supply of Outdoor Recreation: An Econometric Analysis*. New Brunswick, N.J.: Rutgers University Press.

Clark, Colin W. 1976. *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*. New York: J. Wiley & Sons.

Debreu, G. 1959. *The Theory of Value*. New York: J. Wiley & Sons.

Drèze, J., and D. de la Vallée Poussin. 1971. A Tâtonnement Process for Public Goods. *Review of Economic Studies* 37:133-150.

Drèze, J.P., and N.H. Stern. 1987. The Theory of Cost-Benefit Analysis. In *Handbook of Public Economics*, Vol. II, by A. Auerbach and M. Feldstein (eds.). Amsterdam: North-Holland.

Farquharson, R. *Theory of Voting*. New Haven, Conn.: Yale University Press.

Feltham, G.A., A.H. Amershi, and W.T. Ziemba (eds.) 1988. *Economic Analysis of Information and Contracts: Essays in Honour of John E. Butterworth*. Boston: Kluwer Academic Publishers.

Fisher, A.C. 1973. Environmental Externalities and the Arrow-Lind Public Investment Theorem. *American Economic Review* 63:722-725.

Fisher, A.C. 1989. Energy and the Environment in the Long Term. *Energy Policy*. Special Issue 17 (2 April):84-88.

Fisher, A.C., and W.M. Hanemann. 1985. Endangered Species: The Economics of Irreversible Damage. In *Economics of Ecosystem Management*, by D.O. Hall, N. Myers and N.S. Margaris (eds.), 129-138. Dordrecht: W. Junk Publishers.

Fisher, A.C., J.V. Krutilla, and C.J. Cicchetti. 1972a. Alternative Uses of Natural Environments: The Economics of Environmental Modification. In *Natural Environments: Studies in Theoretical and Applied Analysis*, by J.V. Krutilla (ed.). 18-53. Baltimore: Johns Hopkins University Press.

Fisher, A.C., J.V. Krutilla, and C.J. Cicchetti. 1972b. The Economics of Environmental Preservation: A Theoretical and Empirical Analysis. *American Economic Review* 62:605-619.

Fisher, A.C., J.V. Krutilla, and C.J. Cicchetti. 1974. The Economics of Environmental Preservation: Further Discussion. *American Economic Review* 64:1030-1039.

Freeman, A. Myrick III. 1984. The Quasi-Option Value of Irreversible Development. *Journal of Environmental Economics and Management* 11:292-295.

Gallagher, David R., and V. Kerry Smith. 1985. Measuring Values for Environmental Resources under Uncertainty. *Journal of Environmental Economics and Management* 12:132-143.

Gibbard, A. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica* 41:587-602.

Griffin, J. 1977. Are There Incommensurable Values? *Philosophy and Public Affairs* 7:39-59.

Hahn, R.W., and G.L. Hester. 1989. Marketable Permits: Lessons for Theory and Practice. *Ecology Law Quarterly* 16(2):361-406.

Hammond, P.J. 1980. Cost-Benefit as a Planning Procedure. In *Contemporary Economic Analysis*, by D.A. Currie and W.A. Peters (eds.), 221-250. London: Croom-Helm.

Hammond, P.J. 1983. Approximate Measures of the Social Welfare Benefits of Large Projects. Technical Report 459. Stanford: Institute for Mathematical Studies in the Social Sciences.

Hanemann, W. Michael. 1984. *Willingness to Pay and Willingness to Accept: How Much Can They Differ?* Working Paper 328. Berkeley: Department of Agricultural and Resource Economics, University of California.

Hanemann, W. Michael. 1989. Information and the Concept of Option Value. *Journal of Environmental Economics and Management* 16:23-37.

Harvey, A.C. 1981. *Econometric Analysis of Time Series*. Oxford: Philip Allan.

Heal, G. 1973. *The Theory of Economic Planning*. Amsterdam: North-Holland.

Henry, C. 1974a. Investment Decisions Under Uncertainty: The 'Irreversibility Effect'. *American Economic Review* 64:1006-1012.

Henry, C. 1974b. Option Values in the Economics of Irreplaceable Assets. *Review of Economic Studies Symposium* 41:89-104.

Inman, R.P. 1987. Markets, Governments, and the 'New' Political Economy. In *Handbook of Public Economics*, Vol. II, by A.J. Auerbach and M. Feldstein (eds.), 647-777. Amsterdam: North-Holland.

Johansen, L. 1963. Some Notes on the Lindahl Theory of Determination of Public Expenditures. *International Economic Review* 4:346-358.

Johansson, Per-Olov. 1987. *The Economic Theory and Measurement of Environmental Benefits*. Cambridge, England: Cambridge University Press.

Johansson, Per-Olov, and Karl-Gustaf Löfgren. 1985. *The Economics of Forestry and Natural Resources*. Oxford: Basil Blackwell Ltd.

Kay, J. J. 1991. The Concept of Ecological Integrity, Alternative Theories of Ecology, and Implications for Decision-Support Indicators. In *Economic, Ecological, and Decision Theories: Indicators of Ecologically Sustainable Development*, 23-58. Ottawa: Canadian Environmental Advisory Council.

Kennedy, P. 1985. *A Guide to Econometrics*. Cambridge, Mass.: MIT Press.

Knetsch, J., and J. Sinden. 1984. Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. *Quarterly Journal of Economics* 99:507-521.

Knetsch, J., and J. Sinden. 1987. The Persistence of the Valuation Disparity. *Quarterly Journal of Economics* 102:691-695.

Lindahl, E. 1919. Just Taxation – A Positive Solution. Reprinted in *Classics in the Theory of Public Finance*, 1967, by R. Musgrave and A. Peacock (eds.), 168-176. New York: St. Martin's Press.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, England: Cambridge University Press.

Malinvaud, E. 1971. A Planning Approach to the Public Good Problem. *Swedish Journal of Economics* 73:96-112.

Marglin, S.A. 1963. The Social Rate of Discount and the Optimal Rate of Investment. *Quarterly Journal of Economics* 77:95-102.

Markandya, A., and D. Pearce. 1988. *Environmental Considerations and the Choice of Discount Rates in Developing Countries*. Working Paper 3. New York: World Bank, Environment Department.

May, K. 1952. A Set of Independent, Necessary and Sufficient Conditions for Simple Majority Decision. *Econometrica* 20:680-684.

Mueller, D. 1979. *Public Choice*. Cambridge, England: Cambridge University Press.

Norton, Bryan. 1988. Commodity, Amenity, and Morality: The Limits of Quantification in Valuing Biodiversity. In *Biodiversity*, by E.O. Wilson (ed.), 200-205. Washington: National Academy Press.

Nozick, R. 1974. *Anarchy, State and Utopia*. New York: Basic Books.

Partridge, E., ed. 1981. *Responsibilities to Future Generations: Environmental Ethics*. Buffalo, N.Y.: Prometheus.

Paulik, G.J. 1971. Anchovies, Birds, and Fishermen in the Peru Current. In *Environment: Resources, Pollution, and Society*, by W.W. Murdoch (ed.), 156-185. Stamford, Conn.: Sinauer Associates.

Pearce, D., and R.K. Turner. 1990. *Economics of Natural Resources and the Environment*. Hertfordshire: Harvester Wheatsheaf.

Plott, C. 1967. A Notion of Equilibrium and Its Possibility under Majority Rule. *American Economic Review* 57:787-806.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.

Romer, T., and H. Rosenthal. 1979. Bureaucrats vs. Voters: On the Political Economy of Resource Allocation by Direct Democracy. *Quarterly Journal of Economics* 93:562-587.

Romer, T., and H. Rosenthal. 1982. Median Voters or Budget Maximizers: Evidence from School Expenditure Data. *Economic Inquiry* 20:556-578.

Ruitenbeek, H.J., and M. King. 1988. Survey of Utility Planning Capabilities and Reliability. Montreal: Canadian Electrical Association.

Ruitenbeek, H.J. 1990. Evaluating Economic Policies for Promoting Rainforest Conservation in Developing Countries. London: London Stock Exchange.

Samuelson, P.A. 1964. Discussion. *American Economic Review* 54:93-96.

Sandmo, A. 1972. Discount Rates for Public Investment Under Uncertainty. *International Economic Review* 13:187-302.

Satterthwaite, M. 1975. Strategy-Proofness and Arrow's Conditions. *Journal of Economic Theory* 10:187-217.

Scalon, T. Liberty, Contract, and Contribution. In *Markets and Morals*, by G. Dworkin, et al. (eds.), 43-67. Washington: Hemisphere.

Schmalensee, R. 1972. Option, Demand and Consumer's Surplus: Valuing Price Changes under Uncertainty. *American Economic Review* 62:813-824.

Schmalensee, R. 1976. Public Investment Criteria, Insurance Markets, and Income Taxation. *Journal of Public Economics* 5:425-445.

Sen, A. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden-Day.

Sen, A. 1983. Social Choice Theory. In *Handbook of Mathematical Economics*, by K. Arrow and M. Intriligator (eds.). Amsterdam: North-Holland.

Shepsle, K. 1979. Institutional Arrangements and Equilibrium in Multidimensional Voting Models. *American Journal of Political Science* 23:27-59.

Simmons, A.J. 1983. Inalienable Rights and Locke's Treatise. *Philosophy and Public Affairs* 12:175-204.

Smith, V. Kerry. 1983. Option Value: A Conceptual Overview. *Southern Economic Journal* 654-668.

Smith, V. Kerry. 1987. Uncertainty, Benefit-Cost Analysis, and the Treatment of Option Value. *Journal of Environmental Economics and Management* 14:283-292.

Stone, C. 1972. Should Trees Have Standing? Towards Legal Rights for Natural Objects. *Southern California Law Review* 45:450-501.

Vickrey, W. 1960. Utility, Strategy and Social Decision Rules. *Quarterly Journal of Economics* 74:507-535.

Victor, P. A. 1991. Indicators of Sustainable Development: Some Lessons from Capital Theory. In *Economic, Ecological, and Decision Theories: Indicators of Ecologically Sustainable Development*, 1-22. Ottawa: Canadian Environmental Advisory Council.

Viscusi, W.K. 1988. Irreversible Environmental Investments with Uncertain Benefit Levels. *Journal of Environmental Economics and Management* 15:147-157.

WCED [World Commission on Environment and Development]. 1987. *Our Common Future*. London: Oxford University Press.

Weisbrod, B.A. 1964. Collective Consumption Services of Individualized-Consumption Goods. *Quarterly Journal of Economics* 78:471-477.

Wicksell, K. 1896. A New Principle of Just Taxation. Reprinted in *Classics in the Theory of Public Finance*, 1967, by R. Musgrave and A. Peacock (eds.), 72-118. New York: St. Martin's Press.

Willig, R. 1976. Consumer's Surplus without Apology. *American Economic Review* 66:589-597.

Wilson, R. 1972. Social Choice Theory without the Pareto Principle. *Journal of Economic Theory* 5:478-486.

Appendix A

Arrow's Axioms Required of Collective Choice Processes

Pareto optimality (P)	The collective choice process should always choose (denoted by strict preference – *P*) alternative *x* over alternative *y* if everyone in society prefers *x* to *y*. For all individuals *i*, *xPy* if xP_iy
Non-dictatorship (ND)	There is no situation in which one individual's preference is always decisive, even when everyone else prefers the opposite. There is no person $i \mid xP_iy$ and xPy and yP_jx for $j \neq i$
Unrestricted domain (U)	The process is capable of reaching a collective decision on any combination of individual preference orderings of the alternatives.
Rationality (R)	The process must be rational to the extent that: (a) it can express a complete preference between any two alternatives or express indifference (*I*) between them; and (b) all rankings display transitivity of preference. (a) For all *x* and *y*, *xPy* or *yPx* or *xIy* (b) If *xPy* and *yPz*, then *xPz*
Independence of Irrelevant Alternatives (I)	The selection of collective preference of one alternative over another must depend only on the individual's ordering over those two alternatives, and not on individual orderings over other alternatives.

(after Arrow (1963) and Sen (1970))

Appendix B

Attributes of Selected Approaches to Economic Analysis

Analytical Approach	Purpose of Analysis and Attributes of Procedure	Potential Application and Use in Environmental Issues	Information Requirements of Indicators
PREFERENCE DESCRIPTION APPROACHES			
Contingent Value Methods (CVM) **Useful Reference:** Johansson (1987)	ESTIMATE PREFERENCES BY DIRECT SURVEY • Ask people what they are willing to pay (WTP) for a non-market good or service, or what they are willing to accept (WTA) as compensation for loss of a good or service. • Example: Ask people how much they would be willing to pay to have airport noise reduced at night.	*Good Potential Used in Specific Cases* • Can be used to determine how much environmental goods or services are worth to individuals. • Primary constraint is multiple sources of bias from such techniques. One is strategic bias: people will understate value and 'free ride' if they believe that they will actually have to pay; people will overstate value if they believe that they will not have to pay.	*Low* Requires detailed survey of preferences, but can relate only to non-market activities. A potential problem is correcting for discrepancies between WTA and WTP. WTP is what most techniques measure, whereas WTA is the appropriate measure of welfare change. WTP typically understates WTA.
Hedonic Price Methods (HPM) **Useful Reference:** Johansson (1987)	ESTIMATE PREFERENCES BY INFERENCE FROM REVEALED BEHAVIOUR • Observe market prices and infer value of non-market goods after correcting for all potential contributions to observed price. • Usually involves econometric analysis. • Example: Estimate value of noise reduction by comparing property prices near and far from airports or roads.	*Good Potential Used in Specific Cases* • Can be used to determine how much environmental goods or services are worth to individuals. • Primary constraint is analytical difficulty of separating out effects of non-environmental factors.	*Moderate* Requires information relating to only directly observable factors that exist in conventional markets. Information required, however, is extensive as all non-random factors accounting for behaviour must be measured.
Limited Dependent Variable Methods (LIMDEP) **Useful Reference:** Maddala (1983)	ESTIMATE PREFERENCES WHERE DISCRETE VOTING CHOICES ARE AVAILABLE • Usually involves econometric analysis of cross-section data to determine what economic factors affect discrete economic decisions. • Can be component of CVM or HPM. • Example: Observe what makes people take (a) their car or (b) the bus to work.	*Very Good Potential Not Widely Used* • Can be used to predict how people will behave or react. • Greatest potential is in designing economic incentive mechanisms. • Primary constraint to more common usage is that statistical techniques are less well-known than conventional econometric techniques. • Example: Observe what makes a given plant (a) pollute or (b) not pollute.	*Low/Moderate* The major benefit of these methods is that they can handle qualitative information efficiently. Requires enumeration of observed qualitative behaviour for some statistically significant sample. Also requires cross-section of qualitative or quantitative characteristics of sample.

Appendix B (continued)

Attributes of Selected Approaches to Economic Analysis

Analytical Approach	Purpose of Analysis and Attributes of Procedure	Potential Application and Use in Environmental Issues	Information Requirements of Indicators
Structural Description Approaches			
Technical Input-Output (IO-T)	Describe the Physical Flows Between Components of a Physical or Economic System • Usually involves matrix of flows. • Sometimes involves connections to physical stocks. • Usually does not have a common numeraire measure (except with energy balances – calories). • Example: Material balance or energy analysis of physical systems. • Example: Inter-industry flows of goods and services in economy, measured in physical quantities (tons steel, hours labour, etc.).	*Good Potential* *Not Widely Used* • Can be used to describe requirements on environmental resources. • Greatest potential is in connecting these models to models of economic activity (such as IO-EF or IO-EV). • Primary constraint to more common usage is providing connection to economic activities.	*Significant* Requires measurement of physical flows between components of a system. For economic systems, this requires information on the Q component of the V=PxQ identity. If Q is not measured, require prices (P) and gross values of flows (V). A second requirement is measurement of physical stock. For economic systems, this requires information on the K component of the A=PxK identity. If K is not measured, require prices (P) and gross asset value of stock (A).
Economic Input-Output Fixed Coefficients (IO-EF)	Describe the Value of Flows Between Components of an Economic System Without Structural Change • Usually involves matrix of value flows. • Sometimes involves connections to stocks of capital. • Uses money as common numeraire measure. • Assumes constant returns to scale (no structural change). • Example: Value of inter-industry flows of goods and services in economy, measured in dollars (consumer spending, costs of intermediate goods, wages, resource royalties, etc.).	*Good Potential* *Not Widely Used* • Approach is widely used in describing economic flows of marketed goods. • Greatest potential is in integrating environmental goods and services into these models (such as by connecting them to an IO-T model or explicitly adding environmental commodity). • Primary constraint to more common usage is describing physical connection between industries and environmental stocks and flows. • Secondary constraint is providing valuation of environmental stocks and flows.	*Significant* Requires measurement of disaggregated values of flows between components of an economic system. This requires information on the V component of the V=PxQ identity. If V is not measured, require prices (P) and physical flows (Q). A second requirement is measurement of disaggregated values of physical stocks. In some cases, this requires information on the A component of the A=PxK identity. If A is not measured, require prices (P) and amount of physical capital stock (K). As noted by Victor (1990), however, different economic schools will tend to treat the value of capital K differently. Although in some cases it is defined as A=PxK, some economic schools define it as the present value of future income flows.
Economic Input-Output Variable Coefficients (IO-EV)	Describe the Value of Flows Between Components of an Economic System With Structural Change • Usually involves aggregated input values and aggregated output values. • Uses money as common numeraire measure. • Allows for substitution of inputs (structural change). • Example: KLEM models using capital, labour, energy, and materials as input to describe measure of output.	*Good Potential* *Not Widely Used* • Approach is widely applied to individual sectors as well as the economy as the whole. • Greatest potential is in including environmental resources as separate input. • Primary constraint to more common usage is describing physical connection between industries and environmental stocks and flows. • Secondary constraint is providing valuation of environmental stocks and flows.	*Moderate* Similar to IO-EF models described above except they rely on aggregated values. Although the information requirements are less than for IO-EF, the actual aggregation procedure (involving monetization of environmental flows and stocks) can be complex. The same comments associated with capital valuation in IO-EF apply here.

Appendix B (continued)

Attributes of Selected Approaches to Economic Analysis

Analytical Approach	Purpose of Analysis and Attributes of Procedure	Potential Application and Use in Environmental Issues	Information Requirements of Indicators
Impact Description Approaches – Static			
Input-Output (IO)	Predict Impact on Flows from Change in Some Input • Approach involves shocking one of the descriptive structural IO models and predicting the impact of the shock. • Prices are assumed fixed; values and quantities change according to structure of IO model. • Example: By building a hydroelectric dam, we try to predict the impact this has on labour demand, concrete requirements, etc.	*Very Good Potential* *Not Widely Used* • Approach is widely used in describing economic impacts of marketed goods, but rarely extended to include environmental components. • Greatest potential is in linking environmental goods and services into economic IO models. • Primary constraint to more common usage is describing physical connection between industries and environmental stocks and flows.	*Moderate/Significant* Requires a structural IO model with characteristics as described in previous table.
Partial Equilibrium (PE)	Predict Equilibrium Prices and Flows within Individual Market Segment • Approach involves describing or shocking a structural model and predicting prices and flows in one sector. • Prices in other sectors are assumed fixed. • Equilibrium achieved by assuming utility-maximizing agents. • Example: Simple supply and demand models for environmental goods and services.	*Very Good Potential* *Not Widely Used* • Approach is widely used to predict partial equilibrium in describing economic impacts of marketed goods. • Greatest potential is in policy formulation for designing economic incentives relating to environmental goods and services. • Primary constraint is valuation of environmental goods and services.	*Moderate* Can focus on single sector. Requires valuation of environmental goods and services.
General Equilibrium (GE)	Predict Equilibrium Prices and Flows across Market Segments • Approach involves describing or shocking a structural model and predicting prices and flows in all sectors. • Equilibrium achieved by assuming utility-maximizing agents. • Very common to have multiple equilibria (see Figure 2 in text relating to Edgeworth Box).	*Constrained Potential* *Not Widely Used* • Approach is widely used to predict equilibrium in describing economic impacts of marketed goods. • Primary constraint is theoretical: still requires an existence theorem to ensure that stable concurrent equilibria between environmental and economic systems do exist. • Second constraint is practical: requires description of all physical flows.	*Very Significant* Requires description of all material balances and demands. Requires valuation of environmental goods and services in individual preferences. All prices are determined based on material flows. Some deterministic solutions do not require prices of any goods (these are generally fixed coefficient technical structures with constant returns to scale).
Mathematical Programming (MP)	Predict Optimal Flows within Individual Market Segment • Approach involves describing optimal behaviour by assuming utility-maximizing agents. • Prices in all sectors are assumed fixed. • Constraints specified independently. • Approach determines optimal physical or financial flows. • Shadow prices of constraints are calculated.	*Very Good Potential* *Not Widely Used* • Approach is widely used to predict optimal behaviour for conventional marketed goods. • Greatest potential is in linking environmental constraints into optimization models to determine if constraints are binding, and shadow prices of constraints, if they are binding.	*Low/Moderate* Can focus on single sector. Does not require valuation of environmental goods and services, but does require identifying those that are possible constraints, and what the physical demands of those goods are.

Appendix B (continued)

Attributes of Selected Approaches to Economic Analysis

Analytical Approach	Purpose of Analysis and Attributes of Procedure	Potential Application and Use in Environmental Issues	Information Requirements of Indicators
IMPACT DESCRIPTION APPROACHES – DYNAMIC ("FORECASTING")			
Dynamic Optimization: "Optimal Control Theory" (OC) **Useful Reference:** Clark (1976)	PREDICT OPTIMAL BEHAVIOUR WITHIN INDIVIDUAL MARKET SEGMENT ACROSS TIME PERIODS • Approach involves describing or shocking a dynamic structural model and predicting prices and flows in one sector. • Prices in other sectors are assumed fixed. • Equilibrium achieved by maximizing present value function. • Example: By defining cost function for non-renewable energy stock, we predict future prices and depletion pattern to maximize present value profits.	*Fair Potential* *Used in Specific Cases* • Approach is widely used in describing optimal resource depletion. • Primary constraint to more common usage is describing future exogenous inputs (prices, costs, etc.).	*Significant* Can focus on single sector. Requires valuation of demand for environmental goods and services over time.
Statistical Time-Series Analysis **Useful Reference:** Harvey (1981)	PREDICT FUTURE BEHAVIOUR OF AN INDICATOR FROM PAST BEHAVIOUR OF JUST THAT INDICATOR • Approach involves analysing past variations in an indicator to predict future variations of that indicator. • Includes auto-regressive (AR), auto-regressive integrated moving average (ARIMA), and vector auto-regressive (VAR) techniques. • Example: By measuring past energy consumption, we predict future consumption.	*Poor Potential* *Not Used* • Approach is widely used and well developed in forecasting short-term trends. • Primary constraint in environmental applications is that it requires long time-series of data.	*Low* Requires only past values of a given indicator to predict future values. Requires many readings over a long time-series.
Econometric **Useful Reference:** Kennedy (1985)	PREDICT FUTURE BEHAVIOUR OF AN INDICATOR BY FORECASTING EXPLANATORY FACTORS • Approach involves analysing past variations in an indicator by establishing causal links between that indicator and a set of explanatory variables. • Capable of providing structural information, as well as information on substitution and complementarity. • Example: By comparing past energy consumption with past values of income, energy prices, and other prices, we predict future energy consumption after forecasting future values of income, energy prices, and other prices.	*Fair Potential* *Not Widely Used* • Approach is widely used and well developed in forecasting long-term trends. • Primary constraint in environmental applications is that of adequate cross-section or time-series information for performing statistically significant estimates of past or current behaviour.	*Moderate* Can focus on single sector. Requires values of a given indicator, as well as hypothesized causal variables, from past time periods or from cross-section information. Requires forecast of causal factors.
Structural Methods **Useful Reference:** Kennedy (1985)	PREDICT FUTURE BEHAVIOUR OF AN INDICATOR BY FORECASTING FUTURE STRUCTURES • Approach involves predicting what future structure looks like (for example, what the IO characteristics will be) and then forecasting inputs to predict some output. • Example: We predict what the future economic structure looks like (for example, by changing coefficients in an IO model), and then after forecasting some key causal variables for commercial, residential, and industrial activity, we predict energy demand.	*Fair Potential* *Not Widely Used* • Approach is widely used and well developed in forecasting long-term trends. • Primary constraint in environmental applications is establishing a current reference point to which structural changes are made.	*Significant* Usually requires a description of current structures. Requires forecast of causal factors and structural changes.

Appendix B (continued)

Attributes of Selected Approaches to Economic Analysis

Analytical Approach	Purpose of Analysis and Attributes of Procedure	Potential Application and Use in Environmental Issues	Information Requirements of Indicators
IMPACT EVALUATION APPROACHES			
Cost Benefit Analysis (CBA) **Useful Reference:** Drèze and Stern (1987)	IDENTIFY PROJECTS WHICH ARE SOCIALLY BENEFICIAL • Approach involves describing costs and benefits of a given project. • Projects with positive net benefits are beneficial. • Assumes that project itself does not affect shadow prices for valuing goods and services. • Can accommodate distributional equity through explicit welfare-weights. • Example: We evaluate the costs and benefits of a hydroelectric project by considering construction and operating costs, electricity revenues, and any negative and positive externalities such as flood control and wildlife habitat loss.	*Good Potential* *Widely Used* • Approach is widely used and well developed in environmental applications. • Primary constraint is availability of technical flows for environmental goods and services, and valuation of those services.	Moderate Can focus on single project or activity. Requires valuation of both cost and benefit items.
Cost Effectiveness Analysis (CEA) **Useful Reference:** Bradley (1979)	IDENTIFY COSTS OF PROVIDING A GIVEN GOOD OR SERVICE THROUGH A GIVEN PROJECT • Approach involves describing costs of a given project, and identifying a single good or service it is meant to address. • Approach accommodates time series through cost-levelization techniques in a supply price framework. • Primary application is in comparative analyses (between technologies, regions, industries, countries, etc.). • Example: We evaluate costs of different technologies for removing sulphur emissions. Flue gas desulphurization is estimated at $700 per tonne of sulphur removed, fluidized bed combustion is $500 per tonne, etc.	*Very Good Potential* *Used in Specific Cases* • Approach is well developed and sometimes applied in environmental applications for comparative work. • Primary constraint is dealing meaningfully with joint products. • Greatest potential is in evaluating costs of meeting easily defined technical specifications.	*Low/Moderate* Can focus on single project or activity. Requires valuation of only cost items. Requires identification and measurement of flow of single environmental good or service. Requires some benchmark for comparing effectiveness.
Policy Reform Analysis (PRA) **Useful Reference:** Drèze and Stern (1987)	IDENTIFY POLICY REFORMS THAT ARE SOCIALLY BENEFICIAL • Approach involves direct extension of CBA to policy analysis. • Same shadow pricing and valuation rules apply. • Example: Ahmad and Stern (1987) used this for optimal tax design.	*Good Potential* *Not Widely Used* • Approach is well developed but is only recently seeing increased use in conventional policy reform problems. • Primary constraint is estimating impacts of policies on flows of goods and services.	*Moderate/Significant* Can focus on single policy reform. Requires identification of policy impact on flows of goods and services. Requires valuation of both cost and benefit items.
Policy Effectiveness Analysis (PEA)	IDENTIFY COSTS OF PROVIDING A GIVEN GOOD OR SERVICE THROUGH A GIVEN POLICY REFORM • Approach involves direct extension of CEA to policy analysis. • Same shadow pricing and valuation rules apply. • Example: Ruitenbeek (1990) used this for comparison of rainforest conservation policy costs.	*Good Potential* *Not Widely Used* • Approach is well developed but is only recently seeing increased use in conventional policy reform problems. • Primary constraint is estimating impacts of policies on flows of goods and services. • Second constraint is dealing meaningfully with joint products.	Moderate/Significant Can focus on single policy reform. Requires identification of policy impact on flows of goods and services. Requires identification and measurement of flow of single environmental good or service. Requires valuation of only cost items. Requires some benchmark for comparing effectiveness.

Appendix C

Attributes of Risk Analysis Mechanisms

Risk Analysis Approach	Procedure or Examples	Information Requirements of Indicators
Discount Rate Approaches (DR)	Increase discount rate on high-risk projects. Decrease discount rates on projects with long-term impacts.	*None*
Ad Hoc Sensitivity (AHS)	Determine decision under expected values of all inputs and constraints. Change input values on environmentally sensitive inputs to determine their impact on decision.	*Low* Expected values of inputs and constraints.
Critical Value Sensitivity (CVS)	Determine decision under expected values of all inputs and constraints. Determine decision under critical values of all inputs and constraints. OR Determine what 'switching values' of inputs and constraints must be for decision to change. [Examples: gambler's ruin scenarios, worst case scenarios, maximin strategies]	*Moderate* Expected values of inputs and constraints. Specification of some extreme or critical values of input or constraint. OR Single probability of switching values of input or constraint.
Probability-Based Sensitivity (PBS)	Determine decision under acceptable probability value (e.g., 95%) of all inputs and constraints. OR Determine decision under repeated random sampling of all inputs and constraints to generate a decision profile. [Examples: 95% Case; Monte Carlo (MC) simulations; modified MC simulations for correlated inputs]	*Significant* Expected values of inputs and constraints. Probability distribution of input or constraint for single input problems. Conditional probability distributions of inputs and constraints for multiple-input problems.